ASSESSING THE RELATIONSHIPS BETWEEN ONSITE-WASTEWATER TREATMENT SYSTEM MICROBIAL COMMUNITIES, SYSTEM DESIGN, AND ENVIRONMENTAL VARIABLES.

ASSESSING THE RELATIONSHIPS BETWEEN ONSITE-WASTEWATER TREATMENT SYSTEM MICROBIAL COMMUNITIES, SYSTEM DESIGN, AND ENVIRONMENTAL VARIABLES.

By JACOB STUART DEVRIES, B. Sc.

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree of Master of Science

McMaster University MASTER OF SCIENCE (2021) Hamilton, Ontario (Biology)

TITLE: Assessing the Relationships Between Onsite Wastewater Treatment System Microbial Communities, System Design, and Environmental Variables.

AUTHOR: Jacob DeVries

SUPERVISOR: Herb Schellhorn

PAGES: 86

Lay Abstract.

Onsite-wastewater treatments systems such as household septic tanks are vital tools for managing wastewater. However, the microbial ecosystem which digests waste within septic tanks contains unknown interactions that can alter the rate of waste digestion. We used two DNA sequencing methods to assess how microbial communities within septic tanks responded to the tank design and surrounding environment. We then compared results produced by the two sequencing methods. The response of microbial communities to tank design and the environment differed between the two methods. However, the two methods both indicated that one system design produced a more variable microbial community.

Abstract.

Onsite wastewater treatment systems may be improved by altering the design and environmental variables that affect microbial community composition. However, the two most common methods of examining microbial composition through metagenomic sequencing (16S and shotgun sequencing) produce different taxonomic identification results according to microbial community composition and the analytical methods in use. To identify discrepancies between these two sequencing methods, we analyzed the effect of environmental and tank design variables on onsite-wastewater treatment system microbial communities sequenced using both 16S and shotgun sequencing. Shotgun and 16S sequencing produced different results when examining genera-level taxonomic richness, quantifying the effect of system design and environmental variables on community similarity, and identifying differentially abundant taxa between system types. Results were consistent when subjectively examining patterns of community similarity and when examining genera-level taxonomic diversity above 0.1% relative abundance. Identifying methods that produce similar results between 16S and shotgun sequencing supports the reliable analysis of and optimization of OWTS processes.

Acknowledgements

During my thesis program, I have had the opportunity to work with a wide variety of extraordinary people who helped me get where I am now. Since first arriving at McMaster University, each interaction with the administration has left the impression that the needs of students are a priority, and all bureaucratic hurdles were handled with an ease that I have rarely seen from any institution, let alone one the size of McMaster. My supervisor Professor Herb Schellhorn has been an incredible mentor, and I am grateful for all the guidance and patience he has given me while I gradually worked towards creating a thesis and manuscript that I can be proud of as the first step of my career. I have also received invaluable direction from Professor Brian Golding, who gave me a strong foundation in bioinformatics that I hope to continue building on. Professor Jurek Kolasa and Professor Susan Dudley taught me a practical approach to biological statistics that far exceeded my expectations of what I would be able to accomplish in my thesis program. Additionally, the feedback provided by Professors Schellhorn, Kolasa, and Golding in every lab meeting is something that I will continually keep in mind as foundational moments in learning to present information professionally and incorporate criticism into improvements.

I would also like to thank the past and present lab members who have supported me during this program—particularly Alex Chan, James Napthali, Megan Archer, and Faizan Saleem. Even amid chaotic changes and quarantines, everyone I've worked with has been incredibly supportive.

I am also grateful to the volunteers and undergraduate thesis students who put in a tremendous amount of work in Professor Schellhorn's lab. The opportunities to give what little direction I

could were always an exciting opportunity to provide a little of the support that I was given. I will continue to keep in mind the feedback I was given while attempting to teach bioinformatics for the first time.

I would not have reached the end of this program without my wife and family's continual support, all of whom have patiently listened to my ramblings about septic tanks for far too long. I would not have gotten to where I am without the support of so many wonderful people, and I thank you all.

Table of Contents.

List of Figures.

List of Tables.

List of Abbreviations.

ASV: Amplicon Sequence Variant

CAWT: Centre for Advancement of Water and Wastewater Treatment Technologies

COD: Chemical Oxygen Demand

dBg: de Bruijn graph

FAN: Free Ammonia Nitrogen

HVR: Highly Variable Region

NGS: Next Generation Sequencing

NMDS: Non-Metric Dimensional Scaling

OD: Oxygen Demand

OWTS: Onsite Wastewater Treatment System

OTU: Operational Taxonomic Unit

PCR: Polymerase Chain Reaction

PERMANOVA: Permutational Multivariate Analysis Of Variance

TSS: Total Suspended Solids

TKN: Total Kjeldahl Nitrogen

Chapter 1: Review of Metagenomic Methods.

1.1. Introduction.

Historically, the metabolic processes of bacteria were examined within isolated species. However, recent advances in metagenomic sequencing indicate that many metabolic pathways, first studied within isolated species, are probably the results of individual redox reactions occurring across multiple species (Garcia et al., 2018). For example, of 330 denitrifying organisms identified by metagenomic sequencing in 2016, only 12 had genes for the complete denitrification pathway (Anantharaman et al., 2016).

The Interactions between microbial metabolism and biochemical environments can occur at finer scales than are typically examined. Metabolites such as H_2 and CO with low environmental concentrations (under 1mg/l and between 2-17nM, respectively) can remain undetected in chemical analysis (Anantharaman et al., 2016). The involvement of difficult-to-detect biochemical factors in microbial metabolism only becomes apparent when examining the relative abundance of coding sequences involved in specific metabolic pathways (Anantharaman et al., 2016). Due to the lack of sufficient analytical methods for examining microbial metabolism, engineered systems that employ microbial processes have historically been developed using a "black box" approach.

The term "black box" refers more commonly to computing practices in which only the input and output of a given program are identifiable (Nidhra, 2012), but "black box" can also be used to refer to microbial systems in which the intermediate steps of a process are unknown (Manaia et al., 2018). The practical advancement of microbial metabolic tools does not necessarily depend upon knowledge of internal microbial processes but can take the form of advancements in the constructed environment in which they reside. Most elements of bioreactor design are focused on thermodynamic and stoichiometric factors such as heat transfer, agitation, and product removal

(Duan and Shi, 2014; Najafpour, 2015). This approach is exemplified by projects such as the development of a sustained hydrogen-producing bioreactor that receives only electrons as an energy source and only CO_2 as a carbon source (REF) (Jourdin et al., 2015). Pyrosequencing was carried out after the system was established, but its creation depended on establishing the electrochemical conditions that promoted a hydrogen-producing microbial community (Jourdin et al., 2015).

In addition to the design of constructed environments, advancements in the production and study of bioreactors have used bioaugmentation. Bioaugmentation typically involves the addition of biomass taken from a functioning anaerobic digester. The added biomass acts as an enrichment step for the desired but loosely specified community of microbes (Ibrahim et al., 2020; Tale et al., 2015; Venkiteshwaran et al., 2016).

Through either 16S amplicon sequencing or whole metagenome sequencing, metagenomic profiling has been used to gain more specific insights into the microbial community of anaerobic digesters. For example, whole metagenome sequencing of lab-scale anaerobic digesters built to process cow manure has allowed the identification of specific taxa and their associated metabolic processes that drive the anaerobic digestion process (Johnson et al., 2017). The identification of key taxa allows for more accurate monitoring of the waste water treatment design optimization processes (Hassan et al., 2019) and provides insights into potential methods of improving both large and small scale waste treatment systems through the bioaugmentation of a consortium of useful organisms (Chan et al., 2019).

1.2. Sampling.

The exact nature of sampling methods for any study will depend on the environment being sampled. However, identifying the common limitations of sampling can help identify limitations

in the conclusions drawn from later analysis. Microbial community composition can vary spatially on the scale of meters or microns depending on the scales of local environmental conditions (Saetre and Bååth, 2000; Welch et al., 2016). For example, fluorescent in situ microscopy of soil aggregates has indicated that the scale of microbial habitats within this specific environment can range from 2 to 2000 µm (O'Donnell et al., 2007; Raynaud and Nunan, 2014). The environmental conditions of these habitats are determined by biological interactions such as symbiotic microbial relationships that depend on proximity and broader inorganic factors such as temperature or fluid currents. Microbial ecosystems can lose biologically important spatial organization when sampled (Cordero and Datta, 2016).

The compositions of microbial communities may change temporally due to seasonal succession (Jansson and Hofmockel, 2018) or through currently-unidentified processes (Gonze et al., 2018). In terms of anaerobic digesters, varying designs can create temporal and spatial variation through factors such as filters isolating bacterial communities (Herrero and Stuckey, 2014) or chemical gradients emerging as waste moves through a system (Milner et al., 2008).

The modelling of discrete microenvironments allows for the identification of specific scales of microbial habitats that exist in environments of interest. The loss of spatial information during sampling and processing steps can prevent the identification of microbial metabolic interactions that depend on proximity (Cordero and Datta, 2016).

1.3. DNA Extraction.

Sampling, extraction, sequencing, and analysis, all have the potential to introduce artifacts and create inconsistent taxonomic results between studies, potentially obscuring biological trends (Leigh Greathouse et al., 2019). Among these factors, inconsistency in the DNA extraction process contributes the most to varying experimental results (Sinha et al., 2017). DNA extraction can be

performed using standardized commercial kits for the purposes of reproducibility and ease of use (Gaur et al., 2019). These kits are designed to isolate DNA from environments that may contain PCR-inhibiting compounds. These kits also incorporate agents that lyse organisms with high resistance to chemical lysis (for example, archaea which contain S-layers, and Gram-positive bacteria (Roopnarain et al., 2017)). However, there is no kit specifically designed for DNA isolation from anaerobic digesters, despite the presence of PCR inhibiting compounds and high relative abundances of lysis-resistant bacteria and archaea (Roopnarain et al., 2017).

DNA can be isolated from anaerobic digesters using fecal DNA isolation kits (Tuan et al., 2014) or soil DNA isolation kits (Roopnarain et al., 2017). The best practices for isolating DNA from an anaerobic digester include taking steps to prevent phenolic and humic compound contamination (Mahajan et al., 2018; Verma et al., 2017), as well as combining mechanical and chemical lysis for maximum yield of DNA to attain representative species richness and diversity (Roopnarain et al., 2017).

1.4. Next-Generation Sequencing.

Currently, the most common method of DNA sequencing is massively parallel sequencing by synthesis, also referred to as “next-generation sequencing” (NGS). The common processes among methods referred to as NGS include the isolation of short amplicons or fragments of DNA to be sequenced, the preparation of a “library” using annealed sequences (adapters) that bind the target sequences to a surface, and the amplification of library sequences using primers complimentary to the adapter sequences (McCombie et al., 2019). The addition of barcode sequences during the adapters allows target DNA to be labelled according to its sample origin (Chamberlain et al., 1988), allowing multiple samples pooling within one sequencing run and flowcell. This combinatorial process is referred to as multiplexing (Ranjan et al., 2016).

The main features of NGS systems that influence later analytical steps are the number of reads produced and the length of the reads. Illumina systems such as the HiSeq 2500 and MiSeq are among the most commonly used sequencing platforms (Schirmer et al., 2016). The HiSeq 2500 can produce paired-end reads of 2x125bp and create as many as 2 billion reads in slightly under six days. The MiSeq produces up to 25 million paired-end reads with lengths of 2x300bp over about 55 hours (Schirmer et al., 2016).

1.5. 16S Amplicon Sequencing.

The process of 16S amplicon sequencing utilizes high throughput NGS technologies to amplify highly variable regions (HVRs) of the ubiquitous prokaryotic 16S rRNA gene (Chakravorty et al., 2007). During the library preparation phase, one or more of the nine 16S HVRs are amplified from isolated metagenomic DNA. After the amplification of HVRs, a second amplification step is carried out in which adapter and barcode sequences are annealed. Once these amplicons are isolated according to size, sequencing can target the chosen 16S HVRs (Ranjan et al., 2016; Sanschagrin and Yergeau, 2014). The choice of primers and HVRs to be amplified can influence the results of taxonomic analysis (Parada et al., 2016). The V4 HVR using (515f/806r) primers has been shown to yield more similar taxonomic results to whole metagenome sequencing in comparison to V7–V8 (1114f/1392r) and V6–V8 (926f/1392r) primers (Parada et al., 2016). However, there is little consensus regarding HVR and primer choice (Chan et al., 2019). Advancements in long-read third-generation sequencing stand to improve the consistency of 16S taxonomic identifications by using the entire gene (Edgar, 2018; Schloss et al., 2016).

Whether for 16S or whole metagenome sequencing, the processing of Illumina or other NGS data typically begins with the control of common artifacts such as inconsistent quality scores across individual reads, unusually sized fragments, PCR amplification biases, and contamination

(Trivedi et al., 2014). One of the most common tools in NGS sequencing data quality control is **Trimmomatic** (Bolger et al., 2014), a flexible program designed specifically for Illumina sequencing data. This tool processes the detection and removal of technical sequences such as adapters or adapter fragments. **Trimmomatic** also possesses quality filtering functions (Bolger et al., 2014). **BBDuk2** is also a commonly used quality filtering program and is found within the **BBtools** program suite (Bushnell, 2020).

Once quality control of the raw sequences is complete, paired-end reads with overlapping segments can be merged using programs such as **BBMerge** (Bushnell et al., 2017) and **PEAR** (Zhang et al., 2014). Downstream tools may optionally accept sets of unmerged reads. In cases such as 16S rDNA analysis, which utilizes targeted PCR amplification, steps must also be taken to correct pre-sequencing PCR errors, which can create chimeric sequences during the PCR amplification process and lead to the inclusion of false taxonomic groups. Tools such as **ChimeraSlayer** (Haas et al., 2011), **UCHIME2** (Robert C. Edgar, 2016), **DECIPHER** (Wright et al., 2012), and **CATCh** (Mysara et al., 2015) can be used to identify and remove chimeric sequences.

Examining sequencing data both before and after quality control can be accomplished using **fastQC**, a tool for observing aspects of sequencing data, including the read length and average quality scores at given lengths. **FastQC** can be used to identify the effect of processing data using **Trimmomatic** or **BBDuk**, and help to determine which quality control steps are most suited to any given project (Andrews et al., 2015). The results of multiple **fastQC** runs and the results of other processes such as mapping assemblies can be visualized using **MultiQC,** which can create a range of visually appealing and informative graphics (Ewels et al., 2016).

1.6. Taxonomic and Functional Profiling Using 16S Sequencing.

When utilizing 16S rDNA sequencing, targeted HVRs can be clustered into operational taxonomic units (OTUs) based on the similarity between the HVR sequences (Johnson et al., 2019). These clusters can be created using a reference database that adds similar sequences into an OTU or a de novo clustering algorithm that makes pairwise comparisons of sequence similarity (Callahan et al., 2017). The creation of OTUs using HVRs depends on the assumption that sequences with a large percentage of shared identity will have come from closely related taxa (Johnson et al., 2019).

While HVR similarity is reliable for genus-level taxonomic assignment, species-level taxonomic assignments are far less reliable (Johnson et al., 2019). HVRs with a similarity of 97% are commonly considered to be from the same species, and those with 95% similarity are considered to be from the same genus (Chan et al., 2019; Edgar, 2018; Westcott and Schloss, 2015). However, the species level threshold was first established in 1994, when relatively few 16S rDNA sequences were available to verify this threshold (Stackebrandt and Goebel, 1994). More recent proposals suggest that ~99% similarity is necessary for species-level identification using full-length 16S rDNA sequences, and ~100% is necessary when using the V4 hypervariable region (Edgar, 2018).

NGS platforms are currently not capable of sequencing the entire ~1500 bp 16S rRNA gene in a single read, as would be required to use the ~99% species threshold. The 16S rRNA-based phylogenetic analysis tools that use HVR comparison may take the 97% threshold as a given and not allow the user to set a stricter value (Edgar, 2018). Determining whether a tool can set alternate species level thresholds must be done on a case-by-case basis as most common tools undergo frequent updates.

The use of OTUs as identifiers can achieve species-level specificity. However, the process of creating OTUs does not distinguish between true biological variations among HVRs and artificial variations created during the amplification process (Prodan et al., 2020; Rosen et al., 2012). This may lead to false-positive errors in the creation of some OTUs. However, methods that correct amplification errors can reduce the frequency of false-positive errors in the OTU creation process (Callahan et al., 2017). The error-controlled reads, termed amplicon sequence variants (ASVs), represent naturally occurring variations that, unlike OTUs, are consistent between datasets. (Callahan et al., 2017).

Pre-constructed pipelines that contain all or most of the previously mentioned processing steps are most easily differentiated by their use of either OTUs or ASVs. Analytical tools and pipelines such as **MOTHUR** (Schloss et al., 2009), and **USEARCH-UPARSE** (Edgar, 2013), which make use of OTUs, provide less specificity than analytical pipelines such as **USEARCH-UNOISE3** (Robert C Edgar, 2016), **Deblur** (Amir et al., 2017), and **DADA2** (Benjamin J Callahan et al., 2016) which make use of ASVs. These latter pipelines provide greater taxonomic resolution and lower proportions of spurious reads (Prodan et al., 2020).

While 16S sequencing does not provide any direct data regarding the functional genes present in an environment, tools such as **PICRUSt** (Langille et al., 2013) and **Tax4Fun** (Aßhauer et al., 2015) can create approximations of functional gene content. This is done using databases of assembled and genomes with annotated functional genes. **Tax4Fun** (Aßhauer et al., 2015) utilizes the **KEGG** database (Kanehisa, 2004) and **PICRUSt** (Langille et al., 2013) makes use of the **IMG** database (Markowitz et al., 2012). **PICRUSt** also incorporates an OTU-based prediction of gene content (Langille et al., 2013).

1.7. 16S Sequencing in Comparison to Shotgun Sequencing.

Preparatory procedures employed for Shotgun metagenomic sequencing are similar to those of amplicon sequencing. However, instead of amplifying a specific sequence, the first step in library preparation is replaced by the random fragmentation of all DNA present in a sample. This process creates fragments of a selected length suitable for amplification by the chosen NGS sequencing platform (Jovel et al., 2016). Shotgun sequencing increases the taxonomic resolution available, detecting more diversity from highly complex microbial communities, including eukaryotic sequences (Jovel et al., 2016; Ranjan et al., 2016). (Jovel et al., 2016). Shotgun sequencing can also be used to detect the presence of non-prokaryotic organisms without explicitly targeting them for amplification (Breitwieser et al., 2018).

Shotgun and 16S sequencing produce comparable relative abundance data, which decreases in similarity as microbial community complexity increases (Jovel et al., 2016). However, a recent examination of 16S and shotgun metagenomic analysis that used low diversity (15 species) simulated metagenomic datasets found that shotgun sequencing produced a more consistent and accurate assessment of community composition and relative abundance (Khachatryan et al., 2020).

The sequencing of all available DNA allows for more applications than 16S amplicon sequencing (Almeida and De Martinis, 2018; Jovel et al., 2016). The potential metabolic functions of a microbial ecosystem can be assessed using 16S based taxonomic profiling and comparison to the known metabolic capacity of fully-sequenced organisms using databases such as **KEGG** (Aßhauer et al., 2015). However, this process is inherently limited to the assessment of already catalogued metabolic functions.

Shotgun sequencing allows for direct analysis of functional genes present in an environment and can identify novel genes. Identifying novel genes can provide insights into

microbial metabolic functions using homology-based analysis and genes of known function. However, the assembly of novel genes and genomes from shotgun sequencing requires a much greater sequencing depth than is regularly associated with 16S sequencing (Anantharaman et al., 2016; Chan et al., 2019).

When reviewed for comparison (Chan et al., 2019), four different 16S community analysis studies were found to use between $5*10^4$ (Linz et al., 2017) to $1.9*10^5$ (Brown et al., 2017) sequences for analysis, while four shotgun metagenomic studies used between $9.4*10^6$ (Mohiuddin et al., 2017) to $1*10^8$ (Vanwonterghem et al., 2016a) sequences. The only one of those studies to assemble microbial genomes used a read depth of $1*10^8$ (Vanwonterghem et al., 2016a).

The minimum required number of reads for the characterization of a microbial community depends, on the complexity of the community, the method of analysis, and the goals of any given study. Characterizing dominant (>1% relative abundance) bacterial species using 16S and OTU-based methods can take only thousands of reads (Ni et al., 2017), and relative abundances remain consistent when using between 1000 and $5*10^4$ reads (Jovel et al., 2016). However, successful identification of rare species found in fecal communities using 16S V4-V5 HVR sequencing on a MiSeq device requires a minimum of approximately $6.9*10^5$ reads, and $9.82*10^{23}$ reads are required for detailed characterization (Ni et al., 2017).

The range of applications available when using shotgun sequencing data makes estimating the reads required for any given study a case-by-case process. The best practices typically use the results of previous studies which examined the same environments and taxa of interest using the same sequencing platform and analytical methods. MiSeq shotgun sequencing species assignments within six human fecal samples have been observed to level off around 1.7 million reads (Clooney

et al., 2016). However, HiSeq reads can be used to identify additional species past 25 million reads, likely due to the shorter reads produced by HiSeq systems (Clooney et al., 2016).

Assessing both the community composition and the presence of target antimicrobial resistance genes in cattle fecal samples requires a depth of 470 million reads (Zaheer et al., 2018). The rarefaction curve of lower-level taxonomic assignments in cattle fecal samples approaches a plateau at 470 million reads, but additional low abundance (1-6 reads) microorganisms can be identified using 940 million reads (Zaheer et al., 2018). In a study of 33 groundwater samples from a single site and a combination of 4.58 billion reads, genome assembly and metabolic analysis are possible for up to 36% (1,297) of the distinct microorganisms present (Anantharaman et al., 2016).

1.8. Metagenomic Assembly.

Longer sequences must be reconstructed from fragmented sequence reads to assess metabolic functions using NGS shotgun sequencing. Sequenced fragments must be either aligned to a reference genome or assembled, "de novo," using only the sequenced fragments. Due to the short length of individual sequenced fragments, aligning paired-end sequences to a reference genome is more reliable for producing longer contiguous sequences (contigs) (Schirmer et al., 2016). However, reference genomes are not typically available for metagenomic analysis, necessitating more complex de novo assembly methods. The use of de novo assembly methods in a metagenomic context has been facilitated by a decrease in the cost of sequencing, increasing the number of metagenomic sequences available for analysis, and therefore increasing the likelihood of assembling a complete microbial genome (Anantharaman et al., 2016; Wilkins et al., 2019; Wohlgemuth et al., 2018).

Choosing an assembly method depends entirely on the goals of any given project (Quince et al., 2017). Metagenomic genome assembly follows a similar process to isolated single-species

genome assembly (Quince et al., 2017). The most common method for genomic and metagenome assemblies is the de Bruijn graph (dBg) approach. This approach involves breaking reads into overlapping and unique subsequences of a fixed length called k-mers (Pevzner et al., 2001). Together these represent every possible k-length subsequence of the read. Each individual k-mer created represents a vertex (or node) in a dBg, and each node forms a connection (or edge) where the strings of nucleotides overlap by k-1, regardless of the reads which nodes (k-mers) originated.

These connections form what are referred to as edges in the dBg. In the ideal case, this would form a line of overlapping k-mers from one end of the chromosome to another. However, real data containing structural variants, errors, and repeating DNA creates highly complex graphical structures (Pevzner et al., 2001). The variety of dBg based assembly programs represents different heuristic processes for finding the most likely sequence (or path) of overlapping k-mers within the graph (Ayling et al., 2019; Simpson and Pop, 2015). The overlap-layout consensus (OLC) assembly is another available approach, although this method is better suited for assembling longer reads such as those produced by Sanger sequencing or third-generation sequencing (Ayling et al., 2019).

A common method for the assessment of assembly quality is the N50. The N50 represents the minimum contig length which contains 50% of all assembled bases. Tools can be assessed by the trade-off they make between contiguity (N50) and the amount of species and strain diversity represented in a set of assembled contigs (Ayling et al., 2019). The main parameter influencing this trade-off is k-mer size. Decreasing the size of k-mers increases the potential for assembling low abundance genomes, and increasing the size of k-mers decreases the frequency of repeating k-mer sequences, improving the accuracy and length of genome construction (Li et al., 2015; Quince et al., 2017).

Assembly tools such as **MEGAHIT** (Li et al., 2015), **MegaGTA** (Li et al., 2017), and **IDBA-UD** (Peng et al., 2012) attempt to circumvent this trade-off by using an iterative approach involving multiple k-mer sizes. Other tools such as **MetaVelvet-SL** (Afiahayati et al., 2015) separate the collected metagenomic reads into de Bruijn sub-graphs based on differences in k-mer coverage and initial dBg connections. This approach attempts to create sub graphs for individual species, which can then be assembled using single species assembly methods (Afiahayati et al., 2015). The program **metaSPAdes** is a dBg based program which makes use of multiple k-mer sizes and methods originally designed to manage single species genome assembly. **SPAdes** creates assemblies in the presence of nonuniform coverage produced during single cell sequencing (Nurk et al., 2017). The **metaSPAdes** tool makes use of similar methods to manage the nonuniform coverage of metagenomic datasets (Nurk et al., 2017).

Other assembly programs are built for more specific applications or available data types. For example, the **Plass** assembly program can improve the likelihood of assembling low abundance genomes when the focus of a project is the identification of protein sequences. It does so by substituting nucleotides with amino acids and predicted open reading frames for assembly using overlap-based methods. This does not provide an assembled nucleotide sequence but can potentially create more extended amino acid assemblies suitable for identifying novel or taxonomically indicative proteins (Steinegger et al., 2019). In cases where both shotgun sequencing data and long-read sequence data are available, the **OPERA-MS** assembly program can use both these data types to provide more accurate and less fragmented assemblies (Bertrand et al., 2019).

1.9. Taxonomic Profiling Using Shotgun Sequencing.

Taxonomic profiling does not necessarily require the reconstruction of entire or partial genomes. Taxonomic profiles can be created by aligning either assembled contigs or raw reads to databases of taxonomically catalogued sequences. This method is limited by the availability of reference sequences for comparison and alignment. However, non-assembly-based taxonomic profiling can help identify the presence of specific well-catalogued taxa, specifically when the abundances of these taxa are too low for robust assembly (Segata et al., 2012). Programs such as **mOTUs2** (Milanese et al., 2019) and **MetaPhlAn2** (Truong et al., 2015) can provide taxonomic profiles based on sets of taxonomically-indicative marker genes. The mOTUs2 program uses ubiquitous genes to create OTU-based taxonomic classifications (Milanese et al., 2019), and **MetaPhlAn2** uses taxa-specific genes to identify previously-classified taxonomic groups (Truong et al., 2015).

The program Kraken takes a different approach by aligning k-mers of sequenced reads to a dataset containing k-mers and the last common ancestor associated with organisms that contain that k-mer (Wood and Salzberg, 2014). Alignment of protein sequences to a broad array of taxonomically and functionally annotated proteins in the NCBI non-redundant protein catalogue can be accomplished using the **DIAMOND BLAST** tool (Buchfink et al., 2014).

1.10. Functional Profiling Using Shotgun Sequencing.

The proteins within a genome can be identified and annotated using either reference-based alignment programs such as **DIAMOND** or de-novo assembly followed by protein prediction based on assembled open reading frames. Tools such as **MetaGeneMark** (Zhu et al., 2010), **Prokka** (Seemann, 2014), and **Prodigal** (Hyatt et al., 2010) can be used to identify coding sequences and predict proteins from metagenomic assemblies.

The assessment of functional groups and preceding steps can also be accomplished using pipelines that include multiple methods in one software package. The pipelines **MOCAT2** (Kultima et al., 2016) and **MEGAN6** (Huson et al., 2016) can be used for assembly, gene annotation, taxonomic assessment, and functional profiling based on sequence similarity to proteins of known function. **MEGAN6** community edition makes use of **SEED** (Overbeek et al., 2014), **EggNOG** (Huerta-Cepas et al., 2019), **KEGG** (Kanehisa, 2004), and their novel **InterPro2GO** profiling system (Huson et al., 2016). **MOCAT2** utilizes 18 databases, including those used by **MEGAN6**, although they do not include a system built specifically for **MOCAT2** (Kultima et al., 2016).

For a summary of all programs and pipelines discussed here, see table 1.

Chapter 2: Introduction to Anaerobic Wastewater Treatment.

2.1. Impacts of Improperly Treated Wastewater.

The release of improperly treated wastewater into the environment poses a risk to human health and the local ecosystem (Richards et al., 2017). Wastewater from underperforming or failing onsite-wastewater treatment systems (OWTSs) contributes to fecal contamination in recreational and drinking water (Appling et al., 2013; Schaider et al., 2017). This fecal contamination can lead to the spread of waterborne diseases, which are of particular concern in underdeveloped regions where wastewater treatment systems are unavailable or poorly maintained (Adeyemi et al., 2019; Palamuleni, 2002).

Fecal contamination and waterborne diseases are also a concern in regions with relatively well-developed wastewater management infrastructure (DeFlorio-Barker et al., 2018). For example, the United States records approximately 90 million illnesses annually caused by contaminated recreational water (DeFlorio-Barker et al., 2018). Furthermore, the prevalence of

waterborne diseases from contaminated water sources will increase with political instability created by climate change and the emergence of multidrug-resistant pathogens (Caminade et al., 2019).

In addition to harmful microorganisms, poorly treated wastewater contains high levels of nitrogen and phosphorus (Lapointe et al., 2017). These nutrients encourage the growth of algal blooms (Grattan et al., 2016), which can produce toxins that pose a risk to both human health and the local environment. Human exposure to algal toxins can occur either through consuming contaminated seafood or inhalation of aerosolized toxins (Grattan et al., 2016). Environmental damages can occur both through direct harm of marine fauna by algal toxins or as a result of algal blooms decomposing (Karlson et al., 2021). As large algal blooms die off, their decomposition by aerobic microbes creates hypoxic conditions (Karlson et al., 2021). Hypoxic conditions and algae-derived toxins can cause marine habitat loss and severe wildlife mortalities (Al-Yamani et al., 2020; Mallin and McIver, 2012).

2.2. Household OWTS and Anaerobic Digestion.

Household septic tanks are the most common type of OWTS used for managing wastewater without access to a large-scale sewage network (Withers et al., 2014). The most common OWTS design consists of a buried chamber separated into two parts by a wall which wastewater passes over while waste settles in the first chamber (Jowett et al., 2017). In 2017, approximately 1.5 billion people depended on OWTSs for wastewater management (World Health Organization and UNICEF, 2017). Reliance on OWTSs is most common in the urban and periurban regions of middle to low-income countries (Reymond et al., 2018; Sotelo et al., 2019). Despite their abundance and the risks presented by insufficiently treated OWTS runoff (Adeyemi et al., 2019), the typical OWTS design has not significantly changed since the 1800s (Jowett et al., 2017).

OWTSs operate through the anaerobic digestion of organic wastes to reduce the risk of fecal pathogens or nutrient-rich runoff contaminating nearby water sources (Richards et al., 2017). The anaerobic digestion process consists of four basic steps: hydrolysis, acidogenesis, acetogenesis, and methanogenesis (Batstone et al., 2002; Wirth et al., 2012) (see figure 1).

During hydrolysis, acidogenic and acetogenic bacteria release extracellular enzymes, which break down complex organic particulates into soluble monomers that bacteria can utilize (Y. Li et al., 2019). These monomers include simple sugars, fatty acids, and amino acids (Y. Li et al., 2019). The most abundant phyla associated with hydrolytic activity are Fibobacteres, Firmicutes, Bacteroidetes, Spirochaetes, Proteobacteria and Actinobacteria (Qi et al., 2019; Vanwonterghem et al., 2016b).

During acidogenesis, acidogenic bacteria ferment hydrolytic products to produce volatile fatty acids such as propionate, acetate, and butyrate, as well as ethanol (Anukam et al., 2019). The rate of volatile fatty acid production versus ethanol production varies according to pH (Li et al., 2020). Volatile fatty acid production is promoted at a pH of 5 to 6, while a pH of 4 is more conducive to ethanol production (Wu et al., 2017). Increased ethanol production is favourable for the digestion of waste compared to butyric acid or propionic acid (Shi et al., 2021). The production of ethanol is beneficial due to the increased production of hydrogen during acidogenic ethanol production and a more kinetically favourable conversion of ethanol to acetate during acetogenesis (Li et al., 2020; Z. Li et al., 2019; Wu et al., 2017).

The majority of bacteria within anaerobic digestion systems possess some capacity for acetate production (Vanwonterghem et al., 2016b). The most abundant phyla associated with propionate and butyrate production are Actinobacteria, Bacteroidetes, Rhodospirillum, and Verrucomicrobia (Vanwonterghem et al., 2016b). Ethanol-producing bacteria are most common

within the Firmicutes and Proteobacteria phyla (Binod et al., 2013). Ethanol-producing fungi also contribute to the anaerobic digestion process, particularly when digesting lignocellulose-rich waste (Wei, 2016). The most effective fungi at digesting lignocellulose waste are a polyphyletic group referred to as white-rot fungi (Wei, 2016). However, ethanol-producing fungi in anaerobic digestion are most common within the Ascomycota phylum (Binod et al., 2013).

Acetogenesis is defined by acetate production from hydrolytic and acidogenic products (Schuchmann and Müller, 2016). Acetate production can occur through the autotrophic Wood-Ljungdahl pathway (Ljungdahl, 1986; Mayer and Weuster-Botz, 2017) or through a wide range of heterotrophic pathways (Karekar et al., 2019). While many bacteria can produce acetate, the acetate production by bacteria containing the Wood-Ljungdahl pathway differentiates the acetogenic phase of anaerobic digestion from acetate production during the acidogenic phase (Schuchmann and Müller, 2016). Acetogenic bacteria containing this pathway are referred to here as Wood-Ljungdahl acetogens.

Autotrophic acetogenesis relies on the Wood-Ljungdahl pathway to reduce CO_2 to acetyl-CoA using H_2 as an electron donor (Schuchmann and Müller, 2016). Many Wood-Ljungdahl acetogens can also perform heterotrophic acetogenesis, which can use a wide variety of electron donors, including products from both the hydrolytic and acidogenic stages of anaerobic digestion (Karekar et al., 2019). Bacteria containing the Wood-Ljungdahl pathway are most prevalent within the Firmicutes phylum (Schuchmann and Müller, 2016).

Acetate can then be reduced to methane by methanogenic archaea or oxidized by bacteria to produce H_2 and CO_2 (Sun et al., 2014). Acetate oxidizing bacteria rely on syntrophic relationships with methanogenic archaea to remove H_2 and maintain a kinetically favourable

conversion of acetate to H_2 (Westerholm et al., 2019). Acetate oxidizing bacteria are most common in the Clostridia class of the Firmicutes phylum (Westerholm et al., 2019).

Methanogenesis is carried out exclusively by archaea, which typically utilizes H_2 and CO_2, acetate, or methylated compounds (Kurth et al., 2020). These three pathways for methane production are referred to respectively as hydrogenotrophic, acetoclastic, and methylotrophic methanogenesis (Wang et al., 2018). Methanogenesis can also occur using additional or alternative substrates such as ethanol, formate, propanol, butanol, and iron (Kurth et al., 2020). However, these pathways are not as common or as well studied as the hydrogenotrophic, acetoclastic, and methylotrophic pathways (Kurth et al., 2020).

Acetoclastic and hydrogenotrophic methanogenesis are responsible for the majority of global biological methane production (Connelly et al., 2017). Acetate oxidizing bacteria can increase the rate of hydrogenotrophic methanogenesis vs acetoclastic methanogenesis by converting acetate to H_2 and CO_2 (Westerholm et al., 2016). In addition, hydrogenotrophic methanogens and acetate oxidizing bacteria have a higher tolerance for Free Ammonia Nitrogen (FAN) than acetoclastic methanogens (Westerholm et al., 2016). Therefore, high ammonia nitrogen creates an environment that encourages hydrogenotrophic methanogenesis supported by syntrophic ammonia oxidization (Westerholm et al., 2019). The pH of an environment can also influence the relative abundance of hydrogenotrophic and acetoclastic methanogens, as hydrogenotrophic methanogens are more tolerant to acidic conditions (C. Wang et al., 2020).

Furthermore, the autotrophic acetogenesis pathway competes for CO_2 and H_2 with hydrogenotrophic methanogenesis (Boyd et al., 2020). The relative activity of autotrophic acetogenesis and hydrogenotrophic methanogenesis responds to temperature (Fu et al., 2019).

Autotrophic acetogenesis is favoured in low (15ºC) and high temperature (50ºC) conditions, whereas moderate temperatures (30ºC) favour hydrogenotrophic methanogenesis (Fu et al., 2019).

2.3. Inhibition of Anaerobic Digestion and Conditions for OWTS Failure.

Of the four steps of anaerobic digestion, the rate-limiting step of the anaerobic digestion process varies depending on environmental conditions and waste composition (Ma et al., 2013). (Capson-Tojo et al., 2018a). The accumulation of volatile fatty acids such as propionate and acetate is one of the primary indicators of system failure (He et al., 2017). The overaccumulation of volatile fatty acids leads to toxic concentrations of volatile fatty acids and an unfavourably acidic environment for anaerobic digestion (C. Wang et al., 2020). Sudden declines in pH can kill off populations of acetoclastic methanogens, which typically require a pH between 6.8 and 7.2 (C. Wang et al., 2020; Yuan and Zhu, 2016). Mixed populations of acetoclastic and hydrogenotrophic methanogens that have acclimated to gradual pH changes can maintain reactor function at a pH of 4.8 to 5.5. however, without specific steps to cultivate an acid-tolerant microbial community, volatile fatty acid accumulation is likely to prevent adequate OWTS function (Li et al., 2018; C. Wang et al., 2020).

During hydrolysis, the breakdown of high-protein waste can lead to the accumulation of FAN (Yenigün and Demirel, 2013). While beneficial to bacterial growth at low concentrations, FAN can also lead to system failure at higher concentrations of 1 g/L to 1.5 g/L (Capson-Tojo et al., 2020). The archaea responsible for methanogenesis are particularly susceptible to high concentrations of FAN (Yan et al., 2020). Therefore, high ammonia conditions cause system failure through the accumulation of volatile fatty acids and other intermediate compounds due to the decreased rate of methanogenesis (Yan et al., 2020).

The accumulation of volatile fatty acids can also occur if an anaerobic digestion system receives too much high-carbohydrate waste (Yuan and Zhu, 2016). Overloading an anaerobic digestion system with waste that can easily pass through the hydrolysis, acidogenesis, and acetogenesis phases can create an accumulation of volatile fatty acids, leading to system failure (Yuan and Zhu, 2016).

2.4. Prediction of Anaerobic Digester Function.

The microorganisms that determine anaerobic digestion rate live in complex syntrophic relationships, which vary according to environmental variables such as pH, temperature, and waste composition (Capson-Tojo et al., 2018b; Ziels et al., 2016). In addition to the pathways and interactions previously described, many unknown syntrophic relationships and environmental responses determine the stable microbial community composition and the rate of anaerobic digestion within any individual system (García-Lozano et al., 2019; Manaia et al., 2018).

The number of unknown interactions limits the accurate prediction of wastewater anaerobic digester function (Batstone et al., 2015). Predictions of anaerobic digestion rates typically use either a mechanistic model, which includes each known interaction (Batstone, 2006; Batstone et al., 2015), or a machine learning-based approach which examines only operational parameters such as temperature and waste composition (De Clercq et al., 2020; L. Wang et al., 2020). The machine learning-based approach, which does not attempt to model internal microbial interactions, produces better results than the more commonly used mechanistic approach (De Clercq et al., 2020; L. Wang et al., 2020). The most recently developed tools for predicting the function of anaerobic digestion systems use machine learning algorithms that examine both operational parameters and genomic information (Long et al., 2021). Whether predictions attempt to model the entire process (Batstone et al., 2015) or use limited information and machine learning (De

Clercq et al., 2020; Long et al., 2021; L. Wang et al., 2020), an improved understanding of the anaerobic digestion process will aid the design of wastewater treatment systems and the prediction of their success.

The microorganisms that determine anaerobic digestion rate live in complex syntrophic relationships, which vary according to environmental variables such as pH, temperature, and waste composition (Capson-Tojo et al., 2018b; Ziels et al., 2016). OWTS design can contribute to an improved rate of anaerobic digestion through the control of variables such as hydraulic retention time and sediment mixing (Ma et al., 2019). Finding the optimum design conditions for anaerobic digestion is complicated by the amount of unknown or understudied syntrophic interactions occurring within OWTSs (Batstone et al., 2015). Improving the function and design of household OWTSs and larger anaerobic digestion systems depends on an improved understanding of the anaerobic digestion process (Batstone et al., 2015). For an additional summary of recent research on anaerobic digestion syntrophic interactions and environmental effects, see table 2.

Chapter 3: Comparison of Household OWTS Metagenomic Analysis Using 16S and Shotgun Sequencing.

3.1. Introduction.

There are two commonly used sequencing methods for examining microbial ecosystems such as those found in anaerobic digestors. These methods are whole metagenome sequencing (referred to here as shotgun sequencing) and 16S amplicon sequencing (Escobar-Zepeda et al., 2018). However, the results produced by each of these methods can differ depending on environment-specific factors such as multiple species with near-identical 16S rDNA sequences (Antony-Babu et al., 2017). Therefore, it is necessary to determine if 16S and shotgun sequencing

of wastewater anaerobic digesters produce equivalent results when examining the environments within wastewater treatment systems.

In this study, we sought to determine which genetic sequencing methods could be relied upon to meet common research objectives in the study of anaerobic digestion. Community composition data produced using 16S and shotgun sequencing data was used to assess the effect of OWTS design features and environmental variables on the microbial community within the OWTS.

The community composition data used in this comparison was created during the thesis projects of James Naphtali and Wing Yip Alexander Chan (Chan, 2020; Naphtali, 2020). The methods described before the statistical analysis of community composition data, including sampling, DNA extraction, sequencing, quality control, and filtering / trimming, were all carried out by James Naphtali and Wing Yip Alexander Chan (Chan, 2020; Naphtali, 2020).

3.2. Methods.

3.2.1. Sampling Site Description.

The OWTSs examined were all residential installations serving single households. Four distinct system types were tested, each consisting of either a single-pass or recirculating flow type and a conventional or *InnerTube*™ internal design (see figure 2). These four system types were labelled single-pass plug flow (S.P.), single-pass conventional (S.C.), recirculating plug flow (R.P.) and recirculating conventional (R.C.). Conventional OWTSs included a holding tank after the two-chambered main tank All OWTSs sampled also included an aerobic biofilter unit after the initial anaerobic unit. However, this secondary aerobic system was not examined, as the focus of this study was the anaerobic digestion process.

OWTSs with recirculating designs contained an inline valve to direct a portion of the effluent from the aerobic biofilter back to the influent point of the anaerobic OWTS. The degree to which each valve is open is listed as a percentage in table 3. The recirculating valves were set by Waterloo Biofilter Systems Inc and could not be controlled. The Hydraulic retention time was calculated using the system volume (L) divided by the flow rate (L/day).

3.2.2. Sampling Procedure.

Sampling took place from September 2018 to January 2019. The samples were collected from OWTSs across Southern and Central Ontario. Six OWTSs were sampled for each design and flow combination, with the exemption of the S.P systems, for which five systems were sampled. Each system was sampled at three points (see figure 2).

Samples were collected using three devices: the tube sampler, the effluent sampler, and the collection vessel (see figure 3). Samples collected using the tube sampler were emptied into the collection vessel for ease of transport before being distributed into sterile sampling containers. The tube sampler, effluent sampler, and collection vessel were washed with commercial bottled water between uses to minimize cross-contamination.

The first sampling point (influent) in the conventional systems was in the first open chamber beneath the influent pipe. Wastewater was collected from the conventional systems at the influent using the tube sampler (see figure 3). In the plug flow systems, the first sampling point (influent) was at the opening of the innertube beneath the influent pipe. Wastewater was collected from the plug flow systems at the influent point by inserting the tube sampler into the opening of the innertube (see figure 3).

The second sampling point (tank) in the conventional OWTSs was in the second portion of the two-chambered tank. Wastewater was collected from the second sampling point of the

conventional systems using the tube sampler (see figure 3). The second sampling point (tank) in the plug flow systems was the opening at the end of the innertube. Wastewater was collected from the second sampling point of the plug flow systems using the tube sampler (see figure 3).

The third sampling point (effluent) of the conventional systems was the effluent holding tank between the main two-chambered tank and the aerobic biofilter. Wastewater was collected from the third sampling point of the conventional systems using the effluent sampler (see figure 3). The third sampling point (effluent) of the plug flow systems was the spray nozzle feeding into the aerobic biofilter. Wastewater was collected from the third sampling point of the plug flow systems by placing the effluent sampler beneath the spray nozzle and allowing wastewater to spray into the effluent sampler.

The wastewater from each sampling point was distributed into two 500ml and two 100ml wastewater sample bottles. The 100ml sampling containers were each pre-loaded with 200 µl of sulfuric acid (H_2SO_4) preservative. These 100ml containers were used in examining ammonia (NH_3), total Kjeldahl nitrogen (TKN), and chemical oxygen demand (COD). During sample collection, the wastewater in the collection vessel was measured for dissolved oxygen (D.O.), pH, and temperature.

Once sampled, one of the 500ml replicates was sent to McMaster University, and the other was sent to the Centre for Advancement of Water and Wastewater Treatment Technologies (CAWT, Fleming College). All samples were shipped in coolers with ice packs. The longest holding time between sampling and chemical testing at CAWT was seven days. All samples sent to McMaster University exempt for the single-pass plug flow samples were stored at -80^{o}C less than 24 hours after sampling. The plug flow single-pass samples from Central Ontario were kept

on ice for up to six days after sampling. For the location, chemical variables, and environmental variables of each wastewater sample, see table 4.

To assess the effect of the transport and storage protocol, six samples from the influent sampling point and six samples from the tank sampling points were taken from a single plug flow recirculating tank. Three samples from each sampling point were stored at 4^{o}C for one day before being frozen at -80^{o}C to emulate the standard transportation and storage protocol. Three samples from each sampling point underwent DNA extraction on the same day as sampling to act as a control. Lab contamination was assessed by carrying out the DNA extraction process on 200 ml of ddH_2O.

One of the two 500ml wastewater samples and both 100ml samples from each sampling site was sent to the Centre for Advancement of Water and Wastewater Treatment Technologies (CAWT). The remaining 500ml from each sampling site was sent to McMaster University. The samples sent to CAWT were examined for total suspended solids (TSS), total Kjeldahl nitrogen (TKN), ammonia (NH_3). The samples sent to McMaster were analyzed for carbonaceous biological oxygen demand (CBOD).

3.2.3. DNA Extraction.

Samples were thawed and then vacuum filtered through 0.22 um sterile cellulose filters (Brown et al., 2015; Mohiuddin et al., 2019). The sample fluid was added until fluids could no longer pass through the filter. The filters were then transferred to microcentrifuge tubes which were pre-loaded with 0.25ul of 0.1mm zirconium beads (Bag et al., 2016) (BioSpec Products, Bartesville, Oklahoma) and stored at -20^{o}C. DNA was extracted from the filters using the Norgen Biotek soil DNA Isolation Plus Kit (Norgen Biotek Corp., Thorold, Ontario). The manufacturer's extraction procedure was followed unless otherwise specified. The filters and zirconium beads

were immersed in the Norgen lysis buffer and agitated using a Bio spec Products Sonibeast Small Sample Cell Disruptor (Bio specProducts Inc., Bartesville, Oklahoma). The bead beating step included three intervals of 40 seconds. Once the bead beating step was completed, the samples were centrifuged, and the supernatants were transferred to sterile microfuge tubes. Proteins were precipitated using the Norgen acidic precipitation solution. Samples were placed on ice for 5 minutes, then centrifuged again. Humic acid contamination was removed using a Norgen extraction additive, and the samples were again centrifuged. The samples were then filtered through a Norgen DNA-binding centrifuge column. The columns were washed twice using an ethanol solution included in the kit and centrifugation. DNA was then eluted from the columns using the Norgen extraction buffer and centrifugation. Once extracted, the DNA was stored at -20ºC. DNA that was to be used for 16S sequencing was quantified using a NanoDro2000 (Thermofischer Scientific,Waltham, Massachusetts), and DNA that was to be used for shotgun sequencing was quantified using a Qubit 2.0 Fluorometer (Thermofischer Scientific, Waltham, Massachusetts). DNA samples were then stored at - 20ºC.

3.2.4. Shotgun Sequencing, Quality Control, and Classification.

DNA samples were sequenced at the Farncombe Sequencing Institute at McMaster University using an Illumina HiSeq 2500 platform. Paired-end libraries were prepared using the NEBNext® Ultra™ II DNA library preparation kit for Illumina (New England Biolabs Inc.) and TruSeq3 paired-end adapters. The intended read length was 150 bp, and the fragment size was 500 bp. Sequencing took place on two separate lanes. The two FASTQ files for each sample were concatenated using the UNIX cat function. FASTQ files were trimmed using **Trimmomatic** (version 0.39) (Bolger et al., 2014). The Phred score cut-off for **Trimmomatic** was set to 33, and sequences were trimmed using a sliding window with three leading and trailing base pairs, a width

of 4 bases, and a minimum quality score of 20. The TruSeq3 paired-end adapters were also remove by **Trimmomatic**,

All reverse reads were removed from the dataset due to the tendency for Illumina sequencing to produce lower quality reverse reads (Gajer et al., 2021), a lack of overlapping reads, and the inability of **DIAMOND-BLASTx** to process paired-end reads (Buchfink et al., 2014). Read quality was assessed using the R package **fastqcr** (Kassambara, 2019). The mean read count after trimming was 5.3 million reads (SD ± 1.6 million).

DIAMOND-BLASTx (Buchfink et al., 2014) was used to align reads against the NCBI non-redundant (nr) protein database. The e-value cut-off was set at $1x10^{-5}$, the maximum number of target sequences to report alignments for was set to 25 per query. All other options for **DIAMOND-BLASTx** were left as the default. The diamond alignment archive (.daa) output was inputted into **MEGAN6** (version 6.18.4) (Huson et al., 2016) for binning and classification. Sequences were binned according to the **MEGAN6** Weighted Lowest Common Ancestor Algorithm. The minimum quality threshold to assign sequences was set to 50, and the e-value threshold was set to 0.01. Taxonomic assignments were created using the NCBI non-redundant protein database as a reference.

3.2.5. 16S Amplification, Sequencing, Quality Control, and Classification.

The 16S V4 hypervariable region was amplified using a two-stage PCR protocol (Herbold et al., 2015). The 1st PCR stage amplified the V4 hypervariable region, and the 2nd stage attached the adapter for Illumina sequencing and indices for dual-index sample multiplexing (Kozich et al., 2013). Indices were taken from the Nextera XT In dex Kit v2 (Illumina, Inc.). A T100 Thermocycler (Bio-Rad Laboratories, Inc.) was used to carry out the PCR reactions. The PCR

input was normalized to 15 ng/ul using double-distilled water and a NanoDrap 2000 (Thermo Fisher Scientific).

The resulting PCR products underwent gel electrophoresis to confirm that the correct size of amplicon was produced. Once the correct size of amplicon was confirmed, 5 ul of each product was pooled for gel extraction. The extraction of target bands made use of the Axygen Gel Documentation System (Corning Inc.). The extraction and clean-up were performed in triplicate. For each extraction, 25 ul of pooled PCR product was run on 1.8% agarose gel. The gels were examined under 302 nm U.V. light to identify target bands. Target bands were then removed and immersed in a guanidinium thiocyanate solution before being dissolved at 50^{o}C.

The PCR clean-up process made us of a NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel GmbH). The manufacturer's instructions were followed. The columns were washed with ethanol solution three times and then dried using a heat block for 1 minute at 70^{o}C. The NucleoSpin buffer solution was used to elute DNA from each column. Triplicate extractions were then pooled for sequencing. The quality of DNA isolation was verified using a NanoDrop 2000 (Thermo Fisher Scientific).

Amplicons were sent to the Farncombe Institute Genomics Facility and sequenced on an Illumina MiSeq platform (Illumina, Inc.). Like the shotgun sequencing library, the 16S sequencing library was prepared using the NEBNext® Ultra™ II DNA library preparation kit for Illumina (New England Biolabs Inc.) and TruSeq3 paired-end adapters. The sequencing was carried out using a paired-end 300 bp sequencing configuration. The resulting sequences were then demultiplexed. Adapters were trimmed using **cutadapt** (version 1.2.1) (Martin, 2011). Sequence quality trimming and filtering, error modelling, and sequence variant assignment were carried out

using the DADA2 R package and the SILVA 132 SSU reference dataset (Benjamin J Callahan et al., 2016).

3.2.6. Statistical Analysis.

Shotgun sequencing identifications with taxonomic levels labelled as "no ranked" by the NCBI database were removed using the **taxonomizr** R package (version 0.8.0). The original taxonomic labels were replaced by the standard taxonomic ranks from kingdom to species to be compatible downstream with the **phyloseq** (version 1.32.0) R package (McMurdie and Holmes, 2012). From this point onward, the taxonomic identifications produced by 16S and shotgun sequencing were treated identically.

Not all sampled OWTSs were sequenced using shotgun sequencing. Twelve OWTSs were sequenced using shotgun sequencing, and 23 systems were sequenced using 16S amplicon sequencing. Samples were matched between the available 16S and shotgun sequencing data. The final dataset included 36 metagenomes, representing 12 systems, each sampled at the influent, tank, and effluent point. These 12 systems represented three systems from each OWTS type (S.P., S.C., R.P., and R.C.).

Both datasets were agglomerated to the genus level to compare identifications at the same taxonomic level. After agglomeration, both datasets were filtered to remove taxa representing less than 0.1% of the total microbial abundance in that sample. This abundance filtering step was included to examine differences in the identification of low-abundance taxa between 16S and shogun sequencing. This step was repeated using thresholds of 0.05% and 0.2% relative abundance to examine the consistency with which taxonomic identifications responded to abundance filtering

The observed species richness of 16S and shotgun sequencing taxonomic identifications were calculated at multiple stages of data treatment (raw, filtered, agglomerated, and agglomerated then filtered) using the **iNEXT** R package (Hsieh et al., 2016) to examine differences in microbial diversity according to sequencing method and data treatment. The **iNEXT** package also extrapolated true species richness from rarefaction curves.

The community composition data produced by 16S and shotgun sequencing were compared using only samples taken from the central portion of the system (see figure 1) and not the influent or effluent points. The effluent point was not examined because of the inclusion of a holding tank in the conventional systems, which may have altered the microbial community composition. The influent point was examined for initial chemical conditions but was not included in microbial community analysis due to a potential lack of digestive activity in newly deposited waste.

To examine patterns of community variation, filtered and agglomerated taxonomic data was normalized using **DESeq2** (version 1.28.1) (Love et al., 2014), and then **phyloseq** was used to create Bray-Curtis dissimilarity matrices (Bray and Curtis, 1957). These Bray-Curtis dissimilarity matrices were used in PERMANOVA testing (Anderson, 2017), which was carried out using the **VEGAN** adonis2 function to determine the significance of sample clustering according to the variables that were independent of OWTS function. These variables were system design (conventional or plug flow), flow type (single-pass or recirculating), temperature, tank volume, flow rate, and chemical oxygen demand at the influent point. The adonis2 function assessed variables non-sequentially. The results of PERMANOVA analysis produced using 16S and shotgun sequencing data were compared to identify discrepancies in the environmental and design variables that were found to be significant.

The same Bray-Curtis dissimilarity matrices used in the PERMANOVA testing were used to create non-metric dimensional scaling (NMDS) ordinations (Jakaitiene et al., 2016; Tzeng et al., 2008) using the **phyloseq** ordinate function to visualize potential clustering patterns created by OWTS design and flow type. The overall pattern of community dissimilarity produced using 16S and shotgun sequencing was compared using Procrustes analysis to quantify the difference between the microbial communities depicted by each sequencing method.

Taxonomic identifications produced using shotgun sequencing and 16S sequencing that were significantly differentially abundant ($P > 0.05$) according to the **DESeq2** Wald test were examined to determine the effect of OWTS design and flow type on specific taxa. The results of differential abundance analysis produced using 16S and shotgun sequencing were compared to assess the consistency with which taxa displayed differential abundance according to system design and flow type. The formula used in **DEseq2** to assess the effect of flow type while controlling for design was (~ Design + Flow), and **DEseq2** formula to assess the effect of design while controlling for flow type was (~ Flow + Design). The functional significance of differentially abundant taxa was also compared between 16S and shotgun sequencing results.

3.3. Results and Discussion.

3.3.1. Taxonomic Richness.

Shotgun sequencing can detect more low abundance taxa than 16S sequencing (Durazzi et al., 2021). However, the differences between 16S and shotgun sequencing taxonomic identification results depend on DNA extraction efficiency and reference genome availability (Tessler et al., 2017). To assess differences in the total identifications made by 16S and shotgun sequencing of OWTS microbiomes, we compared the taxonomic richness depicted by each sequencing method. We repeated this comparison to determine how the differences between 16S

and shotgun sequencing identifications are affected by data treatment steps such as agglomeration to the genus level and filtering out low abundance (0.1%) taxa. The number of identified taxa in each type of OWTS was also compared to identify differences in the trends between system types depicted by 16S and shotgun sequencing. When comparing numbers of total observed genera in each OWTS, single-pass plug flow type reactors displayed the most variation in taxonomic richness between each system (see figure 4). The 16S and shotgun sequencing data included different numbers of identifications but mostly identified the same pattern of relative abundance between each OWTS (see figure 4). Once the genera were filtered to remove identifications below 0.1% relative abundance, the overall numbers of genera identified were similar. However, the pattern of relative taxonomic richness between OWTS sequenced by shotgun sequencing was lost in the filtering step (see figure 5).

For the total and average species and genera count using taxonomic relative abundance cut-offs of 0.05%, 0.1%, and 0.2%, see tables 5 and 6.

The shotgun sequencing taxonomic identification results from 12 OWTS initially included 23,819 distinct identifications with an average of 12,837 identifications per system. There was a 22.6% increase between the average observed species richness and the average true species richness predicted by the iNEXT package. After agglomeration to the genus level, the shotgun sequencing data included 3,045 distinct identifications with an average of 2505 identified genera per sample. After abundance filtering, the shotgun sequencing data included 121 distinct identifications with an average of 120 identifications per sample. After abundance filtering followed by agglomeration, the shotgun sequencing data included 113 distinct identifications, with an average of 113 identifications per sample.

The 16S sequencing taxonomic identification results initially included 3365 distinct ASVs, with an average of 550 distinct ASVs per sample. There was a 1.4% increase between the average observed taxonomic richness and the average true taxonomic richness predicted by the iNEXT package. After agglomeration to the genus level, the 16S sequencing data included 809 distinct ASVs, with an average of 277 distinct ASVs per sample. After abundance filtering, 16S sequencing included distinct 149 ASVs with an average of 88 distinct ASVs per sample. After agglomeration, followed by abundance filtering, the 16S sequencing data included distinct 123 ASVs, with an average of 96 distinct ASVs per sample.

After we agglomerated and filtered the shotgun sequencing and 16S sequencing, there was a remaining total of 193 identified genera, but only 50 genera of those genera appeared in both datasets.

The numbers of taxa that we identified using 16S and shotgun sequencing were different (Shotgun total = 23,819, 16S total = 3365). However, when examining genera with relative abundances higher than 0.1%, the taxonomic richness measured by 16S sequencing was significantly closer to the taxonomic richness measured by shotgun sequencing (Shotgun total = 113, 16S total = 123). While taxonomic richness was much more similar when examining genera above 0.1% relative abundance, only 48 of the 188 total identified genera above 0.1% relative abundance were shared between 16S and shotgun sequencing.

These results indicate that while the taxonomic richness of abundant genera (>0.1% relative abundance) can be estimated using both 16S and shotgun sequencing, these two sequencing methods identify different taxa as above 0.1% relative abundance. The process of agglomerating low abundance shotgun sequencing identifications to the genus level may have led to genera in the shotgun sequencing dataset increasing in relative abundance in comparison to

16S sequencing. The 22.6% difference between observed species richness and predicted true species richness in the shotgun sequencing dataset indicates that low abundance taxa were more abundant in the shotgun sequencing dataset than in the 16S sequencing dataset, which contained only a 1.4% difference between observed and predicted average taxonomic richness. Future analysis of the differences between 16S and shotgun sequencing taxonomic identifications may benefit from a more detailed breakdown of how individual taxa respond to agglomeration and abundance filtering.

3.3.2. Influence of Environment, Tank Design, and Sequencing Method on Observed Community Composition.

Physiochemical variables such as the temperature, system design, and chemical oxygen demand at the influent point can alter the microbial community composition within OWTSs (Zhu et al., 2020). We used PERMANOVA testing to assess the effects of these physicochemical parameters on community similarity between OWTSs.

When examining the effects of physiochemical variables on OWTS community composition, we did not identify any variables as being significant. The available sample size of 12 tanks representing 3 of each tank design and flow type combination was too small to identify the effects of environmental and tank design variables on community dissimilarity using PERMANOVA. However, we gathered additional samples that did not undergo shotgun sequencing. Using a larger sample size, future analysis of this kind may identify significant effects of environment and tank design on community composition.

When examining the patterns of community similarity between OWTSs, non-metric dimensional scaling determined that single-pass plug flow systems displayed the most intergroup and intragroup variation of genus-level community dissimilarity (see figure 6). The second most

variable system type was the single-pass conventional system (see figure 2). Compared to other systems, the increased variation of plug flow single-pass systems may indicate that the plug flow design without recirculated waste may significantly affect the anaerobic digestion microbial community. However, the examination of NMDS ordination plots is subjective. Future ordination-based research on the effect of OWTS design and flow type will benefit from a larger sample size to supplement quantitative PERMANOVA testing with easily visualized patterns of community similarity.

When comparing the patterns of community similarity between 16S and shotgun sequencing, Procrustes analysis indicates that the NMDS ordinations which we created using 16S and shotgun sequencing depicted very similar patterns of community similarity. The symmetric Procrustes analysis identified a sum of squares of 0.0982. The Protest resulted in a significance of 0.001 and a correlation coefficient of 0.950. Therefore, either 16S or shotgun sequencing can be used when the goal of sequencing is to examine the whole-community variation between sampling sites subjectively.

3.3.3. Influence of Tank Design Variables on Specific Taxa.

There are likely many unknown biochemical interactions that influence anaerobic digestion systems' performance (Batstone et al., 2015). However, taxa with known effects on the hydrolytic, acidogenic, acetogenic, or methanogenic stages of anaerobic digestion can consistently impact the rate of waste removal (Zhang et al., 2019). Taxa which consistently correlate with changes in any aspect of the anaerobic digestion process, represent a potential tool for altering the digestor microbiome (Yin et al., 2016; Zhang et al., 2018). To examine the response of specific taxa to tank design and flow type, we used DESeq2 to identify taxa with

average relative abundances that were significantly different between tank designs and flow types.

When comparing plug flow to conventional system designs, we identified two significantly differentially abundant taxa using shotgun sequencing (see table 7) and five significantly differentially abundant taxa using 16S amplicon sequencing (see table 8). We only identified one genus (*Desulfomicrobium*, enriched in conventional systems) as significantly (Shotgun P=2.2E-4, 16S P=5.05E-18) differentially abundant according to both 16S and shotgun sequencing comparing plug flow to conventional system designs. The genus *Desulfomicrobium* (enriched in conventional systems) contains sulphur reducing bacteria, which use oxidized sulphur compounds and elemental sulphur as electron acceptors (Kushkevych, 2013; Sun et al., 2017).

When comparing single-pass systems to recirculating systems, we identified seven significantly differentially abundant taxa using shotgun sequencing (see table 9), and we identified 13 significantly differentially abundant taxa using 16S sequencing (see table 10). When comparing single pass and recirculating flow types, we identified two taxa as significantly differentially abundant according to both 16S and shotgun sequencing, namely *Phenylobacterium* (shotgun P=1.15E-04, 16S P=2.47E-04) and *Simplicispira* (shotgun P=1.15E-04, 16S P=6.97E-04).

Bacteria in the *Phenylobacterium* genus (enriched in single-pass systems) can utilize heterocyclic phenyl compounds such as those found in artificial herbicides and surfactants as carbon sources (Oh and Roh, 2012). The *Phenylobacterium* genus is also associated with cellulose metabolism (Puentes-Téllez and Salles, 2020; Verastigui et al., 2021). Bacteria in the

Simplicispira genus (enriched in single-pass systems) can perform denitrification (Siddiqi et al., 2020).

Phenylobacterium, Desulfomicrobium, and *Simplicispira* improve the function of anaerobic digestion systems (Siddiqi et al., 2020; Sun et al., 2017; Verastigui et al., 2021; Wilhelm et al., 2019).

Given the lack of overlapping differential abundance results between 16S and shotgun sequencing, verifying the taxonomic differential abundances that indicated *Desulfomicrobium, Phenylobacterium,* and *Simplicispira* responded to tank design and flow type is necessary for determining if OWTS design can reliably influence the relative abundance of these potentially useful taxa.

Conclusions.

Despite the widespread use of onsite wastewater treatment systems (OWTSs), and the need to improve their function, there has been limited use of DNA sequencing to examine the microbial communities which drive the anaerobic digestion process within OWTSs. Research into the internal functions of OWTS will depend extensively on both 16S and shotgun sequencing. Therefore, establishing which research objectives can be met reliably with both 16S and shotgun sequencing is essential for developing effective methods in future research.

In this project, we assessed the microbial community composition of Ontario household OWTSs sequenced using both 16S and shotgun sequencing to identify relationships between community composition and environmental and system design variables. Specifically, we assessed the total taxonomic richness across all systems at multiple stages of data treatment, the similarity of each OWTS microbial community to each other OWTS microbial community, and the effect of OWTS design on the relative abundance of specific taxa. We then compared the results of these

analytical steps between 16S and shotgun sequencing to identify which research objectives could be technologies.

We found that the OWTSs designed with a recirculating flow system and plug-flow type design contained the most variable taxonomic richness and that 16S and shotgun sequencing identified similar numbers of taxa when examining genera over 0.1% relative abundance. However, as most taxa identified by shotguns sequencing were below 0.1% relative abundance, the filtration step caused information on the relative taxonomic richness between OWTSs to be lost. The filtered genera level identifications also included mostly different genera between the 16S and shotgun sequencing datasets. Of the 193 genera above 0.1% relative abundance identified, 50 appeared in both the 16S and shotgun sequencing datasets. The difference in identified genera above 0.1% relative abundance was most likely due to the low abundance species in the shotgun sequencing dataset, which increased the abundance of specific genera when we agglomerated all identifications to the genus level. The variable community composition and taxonomic richness observed in single-pass OWTSs may indicate a more pronounced response to environmental conditions and input composition in these OWTSs. However, a larger sample size will be necessary to quantify the effects of environment and system design on OWTSs.

The sample size of 3 replicates per flow type and system design combination may have been too low to quantify the effect of OWTS flow type system design, temperature, flow rate, volume, and chemical oxygen demand at the influent point on microbial community composition. However, we did find that both 16S and shotgun sequencing indicated that the single-pass plug flow type OWTSs contained the most variable microbial community composition when comparing the microbial communities of each OWTS using NMDS ordination. The consistency with which patterns of microbial community similarity were observed using both 16S and shotgun sequencing

provides an opportunity to plan future research with an understanding of which sequencing methods can be reliably used for community similarity comparisons.

While the available sample size was too small to quantify the effects of environment and system design on the entire microbial community, we did successfully quantify the effects of OWTS design variables on specific taxa. The taxa with significantly different relative abundances were different between 16S and shotgun sequencing. Of the 27 genera which we found were significantly differentially abundant, three were significantly differentially abundant according to both 16S and shotgun sequencing. The sulphur-reducing genus *Desulfomicrobium* was enriched in the conventional two-chambered septic tank design, while the denitrifying genus *Simplicispira* and the cellulose and phenyl-degrading genus *Phenylobacterium* were both enriched in OWTSs with single-pass flow systems. While many of the results of differential abundance analysis differed between 16S and shotgun sequencing, the consistent differential abundances of *Desulfomicrobium*, *Simplicispira*, and *Phenylobacterium* indicate that the design variables of OWTS may be used to alter the relative abundances of these genera and potentially improve the anaerobic digestion process. The design-based enrichment of these taxa may support the digestion of particular substrates rich in nitrate, cellulose, sulphur, and phenolic compounds.

In conclusion, the findings of this project provide insight into the effects of OWTS design on community variation and information on the research objectives that can be met using both 16S and shotgun sequencing.

Figures.

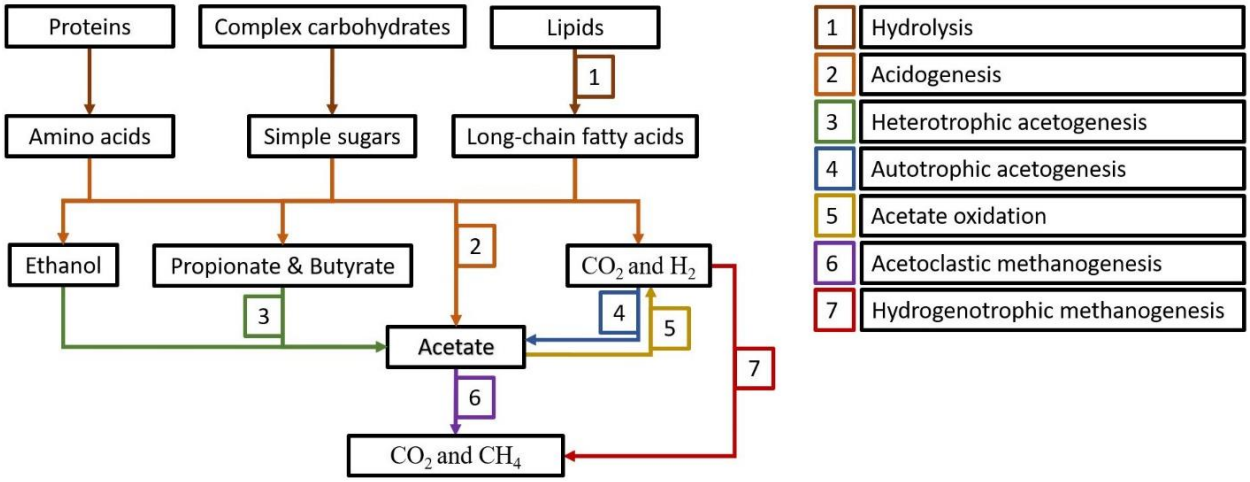
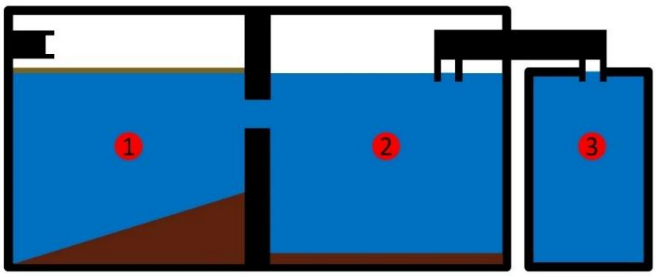


Figure 1. Anaerobic digestion of solid waste components to gaseous products.

A.



B.

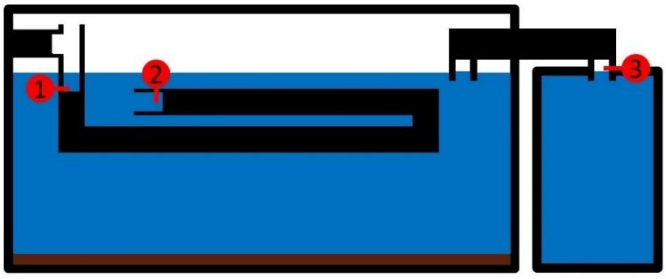


Figure 2. The location of sampling points and internal layouts of OWTSs sampled for metagenomic analysis.

(A) depicts the conventional two-chambered OWTS design with influent sampling site located within the first chamber beneath the influent pipe (1), the tank site located within the second chamber (2), and the effluent site (3) located within the effluent holding tank. (B) depicts the plug flow type InnerTube™ OWTS design with the influent sampling site located at the opening of the innertube (1), the tank site located at the end of the innertube (2), and the effluent site located at the effluent spray nozzle (3).

Figure 3. The effluent sampler, collection vessel, and tube sampler.

The effluent sampler (top) is capped with a tube that has an approximate volume of 0.75 L. The collection vessel (middle) is capped with a tube that has an approximate volume of 1.25 L. The tubes of both the effluent sampler and the collection vessel are affixed by two metallic screws to the bottom of hollow and open-ended poles approximately four feet in length. The tube sampler (bottom) consists of a hollow and open-ended plastic tube approximately four feet in length with a rigid foam sphere at the bottom end (left) attached by a rope knotted at the top end (right). A dual hook is fastened to the top end of the tube (left). Pulling on the rope from the top end (left) seals the tube once it has been inserted into the wastewater to fill.

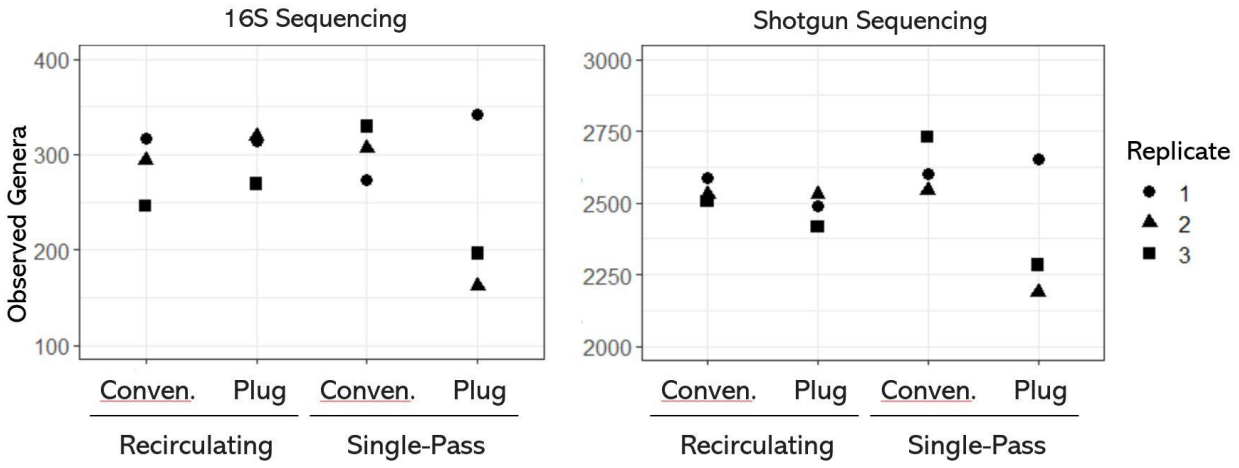


Figure 4. Total genera in each OWTS identified by 16S and shotgun sequencing.

Conventional and plug flow systems labelled Coven., and Plug respectively.

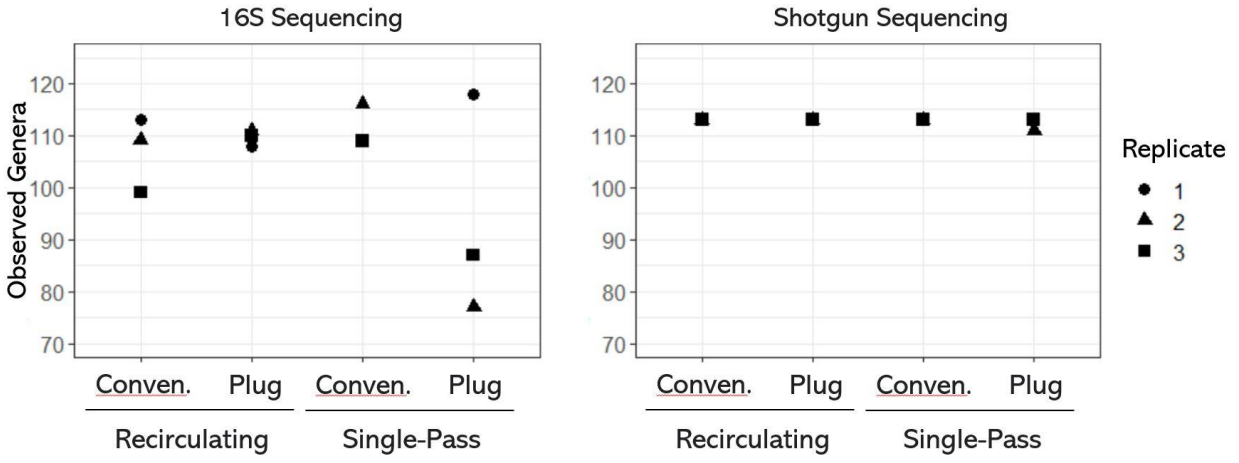


Figure 5. Total genera above 0.1% relative abundance in each OWTS identified by 16S and shotgun sequencing.

Conventional and plug flow systems labelled Conven. and Plug respectively.

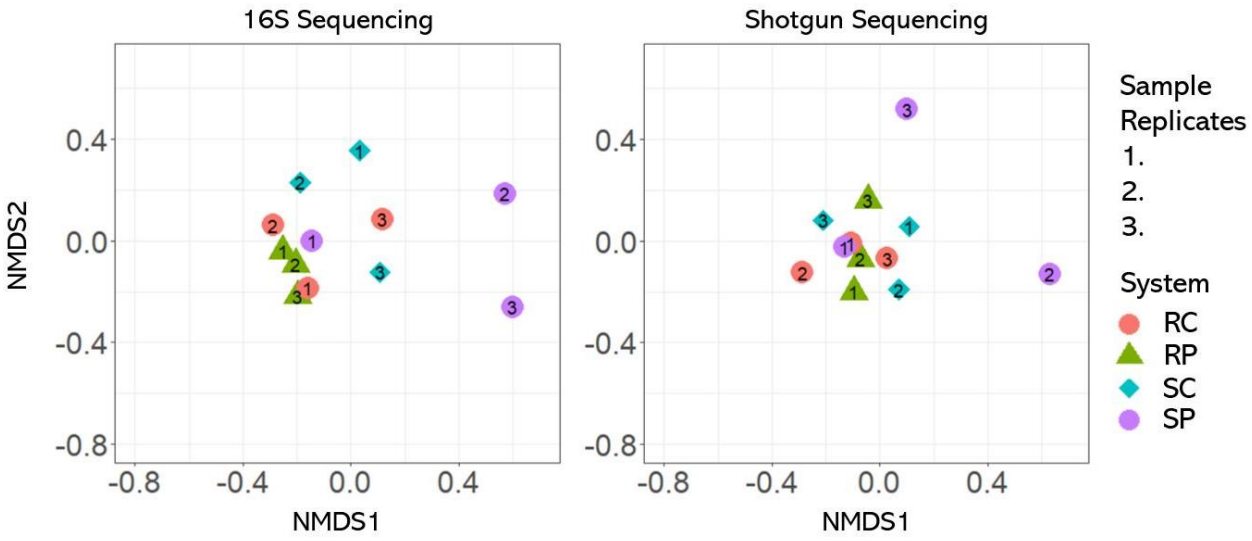


Figure 6. Non-metric dimensional scaling ordination plots of Bray-Curtis dissimilarity matrices of onsite wastewater treatment system microbial communities examined using 16S and shotgun sequencing.

Tables.

Table 1. Primary functions of all programs and pipelines discussed in Chapter 1: Review of Metagenomic Methods.

Program / Pipeline name	Description
Sequence Quality Control	
BBDuk2	A flexible tool within the BBTools program suite for trimming technical sequences and low-quality sections from NGS sequencing data (Bushnell, 2020).
CATCh	A tool for removing chimeric DNA in 16S sequencing data which makes use of both database and de novo chimera identification (Mysara et al., 2015).
ChimeraSlayer	A tool for removing chimeric DNA in 16S sequencing data using a database of chimera-free 16S sequences (Haas et al., 2011).
DECIPHER	A tool for removing chimeric DNA in 16S sequencing data using a search-based approach (Wright et al., 2012).

Program / Pipeline name	Description
fastQC	A quality assessment tool for observing important sequencing data characteristics such as quality scores, length distributions and GC content (Andrews et al., 2015).
MultiQC	A quality assessment tool for visualizing the results of multiple fastQC assessments, as well as the results of later processes such as mapping assemblies (Ewels et al., 2016).
Trimmomatic	A flexible tool which can remove technical sequences and low-quality sections from NGS sequencing data (Bolger et al., 2014).
UCHIME2	A tool for removing chimeric DNA in 16S sequencing data using both a database and de novo based identification (Robert C. Edgar, 2016).
16S Taxonomic and Functional Profiling	
Deblur	An AVS based clustering tool (Amir et al., 2017).
PICRUSt	A functional group prediction tool which uses an OUT based prediction of gene content (Langille et al., 2013), as well as the IMG reference database (Markowitz et al., 2012).
Tax4Fun	A functional group prediction tool which uses the KEGG database (Aßhauer et al., 2015).
USEARCH-UNOISE3	An ASV based clustering tool found in the USEARCH program suite (Robert C Edgar, 2016).
USEARCH-UPARSE	An OTU based clustering tool found in the USEARCH program suite (Edgar, 2013).
16S Analysis Pipelines	
DADA2	An ASV based 16S analytical pipeline which contains steps from quality control to ASV clustering (Benjamin J. Callahan et al., 2016).
MOTHUR	An OTU based 16S analytical pipeline with functions including quality control, OTU clustering, and statistical analysis (Schloss et al., 2009).
Metagenomic Assembly.	
IDBA-UD	A metagenomic assembly tool which uses repeated de Bruijn graphs with multiple k-mer lengths, and it designed to process data with uneven sequencing depth (Peng et al., 2012).
MegaGTA	A gene-targeted metagenomic assembly tool which uses repeated succinct de Bruijn graphs with multiple k-mer lengths (Li et al., 2017).
MEGAHIT	A metagenomic assembly tool which uses repeated succinct de Bruijn graphs with multiple k-mer lengths (Li et al., 2015).
MetaSPAdes	A metagenomic assembly tool which uses repeated de Bruijn graphs with multiple k-mer lengths (Nurk et al., 2017).
MetaVelvet-SL	A metagenomic assembly tool based on decomposing de Bruijn graphs into sub-graphs which can be processed more easily (Afiahayati et al., 2015).
OPERA-MS	A metagenomic assembly tool which can make use of both shotgun sequencing and long read sequencing data (Bertrand et al., 2019).

Program / Pipeline name	Description
Plass	A metagenomic assembly tool which uses amino acid-based assembly of predicted open reading frames (Steinegger et al., 2019).
Taxonomic and Functional Profiling Using Shotgun Sequencing	
DIAMOND	A gene identification tool which aligns protein sequences translated from shotgun sequencing data to the NCBI-nr protein database (Buchfink et al., 2014).
Kraken	A taxonomic profiling tool for shotgun sequencing data which uses a genome database and k-mer alignment to identify the last common *ancestor* which corresponds to the sequence being aligned (Wood and Salzberg, 2014).
MetaPhlAn2	A taxonomic profiling tool for shotgun sequencing data which uses a set of taxonomically specific marker genes to indicate the presence of previously classified taxonomic groups (Truong et al., 2015).
MetaGeneMark	A tool for the de novo prediction of genes from shogun sequencing data (Zhu et al., 2010).
mOTUs2	A taxonomic profiling tool for shotgun sequencing data which uses a set of ubiquitous marker genes to form OTUs (Milanese et al., 2019).
Prokka	A gene identification tool which aligns protein sequences translated from shotgun sequencing data to a set of decreasingly specific databases (Seemann, 2014).
Shotgun sequencing analysis pipelines	
MEGAN6	An analytical pipeline for metagenomic shotgun sequencing analysis (Huson et al., 2016).
MOCAT2	An analytical pipeline for metagenomic shotgun sequencing analysis (Kultima et al., 2016).
Statistical Tools for Metagenomic Analysis.	
DESeq2	A differential abundance analysis tool which includes an advanced normalization process (Love et al., 2014).
Phyloseq	An integrated data analysis tool for taxonomic data and associated data (such as environmental variables) (McMurdie and Holmes, 2012).
taxonomizr	A data handling tool for NCBI taxonomy files and BLAST results (Sherrill-Mix, 2019).
VEGAN	A community ecology tool which includes functions for creating and analysing ordinations (Dixon, 2003).

* Programs / pipelines listed often have additional or secondary functions not listed

Table 2. A summary of recent research on anaerobic digestion syntrophic interactions and environmental effects.

Title	Authors and Year	Journal	Principal finding
Solar Septic Tank: Next Generation Sequencing Reveals Effluent Microbial Community Composition as a Useful Index of System Performance.	(Connelly et al., 2019)	Water (Switzerland)	An increase in temperature driven by solar power correlated to an increased rate of waste removal and an altered microbiome.
Temporal dynamics of activated sludge bacterial communities in two diversity variant full-scale sewage treatment plants.	(Jiang et al., 2018)	Applied Microbiology and Biotechnology	The microbial community within two sewage treatment plants were correlated with influent composition and operating parameters.
Deterministic mechanisms define the long-term anaerobic digestion microbiome and its functionality regardless of the initial microbial community.	(Peces et al., 2018)	Water Research	Long term waste removal performance, metabolic rates, and community composition are independent of starting inoculum.
Influence of Temperature on Biogas Production Efficiency and Microbial Community in a Two-Phase Anaerobic Digestion System.	(Wang et al., 2019)	Water (Switzerland)	Temperatures above 25ºC promoted methanogenesis, and temperatures below 20ºC inhibited methanogenesis and acidogenesis.
Thermophilic anaerobic digestion: Effect of start-up strategies on performance and microbial community.	(Shin et al., 2019)	Science of the Total Environment	During the start-up of wastewater treatment plants, a one-step increase in temperature promoted volatile solid removal, while step-wise temperature change promoted methanogenesis.
Variable sediment methane production in response to different source-associated sewer sediment types and hydrological patterns: Role of the sediment microbiome.	(Chen et al., 2021)	Water Research	Continual hydraulic flow without turbulent suspension of sediments produces more methane from sewer sediments than interrupted flow. Hydrogenotrophic methanogenesis is predominant in storm sewers and illicit discharge-associated storm sewers.

Title	**Authors and Year**	**Journal**	**Principal finding**
Microbiome taxonomic and functional profiles of two domestic sewage treatment systems.	(Hidalgo et al., 2021)	Biodegradation	A school septic system showed lower COD removal rates than a factory septic system with a higher abundance of Chloroflexi, Bacterioidetes, and Proteobacteria.
Addition of granular activated carbon and trace elements to favor volatile fatty acid consumption during anaerobic digestion of food waste.	(Capson-Tojo et al., 2018a)a	Bioresource Technology	The addition of activated carbon to an anaerobic digestion system promoted the growth of archaea and syntrophic bacteria while improving acetate degradation. The addition of trace elements improved the degradation of propionate.
Acclimation of Acid-Tolerant Methanogenic Culture for Bioaugmentation: Strategy Comparison and Microbiome Succession.	(C. Wang et al., 2020)	ACS Omega	A gradual decrease in pH allowed the development of *Methanothrix* and *Methanolinea* populations which support anaerobic digestion, while a sudden decrease in pH led to a halt in methane production.
The fate of anaerobic syntrophy in anaerobic digestion facing propionate and acetate accumulation.	(Yue et al., 2021)	Waste Management	Acetate and propionate negatively influenced biogas production. However, hydrogenotrophic methanogens maintained methane production during acetate and propionate accumulation.
Characterisation of microbial communities for improved management of anaerobic digestion of food waste.	(de Jonge et al., 2020)	Waste Management	Microbial communities in industrial-scale reactors digestion food waste were influenced by substrate type and temperature.
Effect of ammonia on anaerobic digestion of municipal solid waste: Inhibitory performance, bioaugmentation and microbiome functional reconstruction.	(Yan et al., 2020)	Chemical Engineering Journal	The addition of *Methanoculleus* to an anaerobic digestion system improved methane yield through syntrophy with *Peptococcaceae*, *Syntrophaceticus,* and *Peptococcaceae*
Novel insights into the anaerobic digestion of propionate via *Syntrophobacter fumaroxidans* and *Geobacter sulfurreducens*: Process and mechanism.	(Wang et al., 2021)	Water Research	Co-culture of *Syntrophobacter fumaroxidans* and *Geobacter sulfurreducens* improved propionate degradation compared to a co-culture of *Syntrophobacter fumaroxidans* and methanogens.

Title	**Authors and Year**	**Journal**	**Principal finding**
Quantifying the percentage of methane formation via acetoclastic and syntrophic acetate oxidation pathways in anaerobic digesters	(Jiang et al., 2017)	Waste Management	Hydrogenotrophic methanogenesis produced 68% to 75% of the methane from an anaerobic digestion system under high total ammonia nitrogen conditions and produced 9% to 23% of the methane under low total ammonia nitrogen conditions.
Acetoclastic methanogenesis led by Methanosarcina in anaerobic co-digestion of fats, oil and grease for enhanced production of methane	(Kurade et al., 2019)	Bioresource Technology	The co-digestion of sewage waste with fats, oil, and grease led to an increase in acetoclastic methanogenesis led by *Methanosarcina* and a 217% increase in overall methane production.
Mitigation of ammonia inhibition through bioaugmentation with different microorganisms during anaerobic digestion: Selection of strains and reactor performance evaluation	(Yang et al., 2019)	Water Research	The addition of *Methaobrevibacter smithii* and *Syntrophaceticu schinkii* increased methane production of an anaerobic digestion system by 71.1%, while the addition of *Methanosarcina barkeri* alone increased methane production by 59.7%.
Competition Between Chemolithotrophic Acetogenesis and Hydrogenotrophic. Methanogenesis for Exogenous H2/CO2 in Anaerobically Digested Sludge: Impact of Temperature	(Fu et al., 2019)	Frontiers in Microbiology	At temperatures of 15ºC and 50ºC chemilitotrophic acetogenesis utilized more H_2 and CO_2 than hydrogenotrophic methanogenesis, while at 30ºC, H_2 and CO_2 the inverse occurred.
Effect of pH on volatile fatty acid production and the microbial community during anaerobic digestion of Chinese cabbage waste.	(Zhou et al., 2021)	Bioresource Technology	In an anaerobic digestor processing cabbage, acetate, propionate, and butyrate were the primary products while the pH was at 6.0, while acetate, propionate, and butyrate were the primary products at a pH of 7.

Title	Authors and Year	Journal	Principal finding
Enhanced methane production in an anaerobic digestion and microbial electrolysis cell coupled system with co-cultivation of *Geobacter* and *Methanosarcina*.	(Yin et al., 2016)	Journal of Environmental Sciences (China)	The addition of co-cultivation of *Geobacter* and *Methanosarcina* in an anaerobic digestion microbial electrolysis cell resulted in a 24.1% increase in methane production relative to inoculum without the added microbes.
Effect of mixing intensity on hydrolysis and acidification of sewage sludge in two-stage anaerobic digestion: Characteristics of dissolved organic matter and the key microorganisms.	(Ma et al., 2019)	Water Research	In a continuously stirred tank reactor, a mixing speed of 90 and 120 rpm enhanced hydrolysis and acidogenesis.

Table 3. Design variables of septic systems sampled for metagenomic analysis.

System Replicate[1]	Tank Residence Volume (L)	Flow Rate (L/day)	Hydraulic Residence Time (days)	Recirculating Valve Open
1SP	4830	800	6	No Valve
2SP	6060	800	8	No Valve
3SP	3600	1000	4	No Valve
1RP	24000	4000	6	Unknown
2RP	13000	2625	5	50%
3RP	17280	2888	6	10%
1SC	9170	2000	7	0%
2SC	9170	2000	5	No Valve

System Replicate[1]	Tank Residence Volume (L)	Flow Rate (L/day)	Hydraulic Residence Time (days)	Recirculating Valve Open
3SC	5400	1000	5	No Valve
1RC	5400	1250	4	50%
2RC	9173	1750	5	25%
3RC	9170	2250	4	50%

1. System types are RC (recirculating conventional), SC (single-pass conventional), RP (recirculating plug flow), and SP (single-pass plug flow)

Table 4. Location, chemical, and environmental variables of all wastewater samples.

Sample Code[1]	Location	Sample Type	pH	DO (mg/L)	Temp. (ºC)	COD (mg/L)	TSS (mg/L)	TKN (mg/L)	NH3 (mg/L)
1RP1	75 Firelane 2	Inlet	7.42	1.86	16.7	326	138	62.8	54.3
1RP2	75 Firelane 2	Tank	7.47	1.86	17.1	227	52	72.0	65.2
1RP3	75 Firelane 2	Effluent	7.61	5.58	16.9	243	44	72.8	67.0
2RP1	1552 Concession 2	Inlet	7.27	4.84	19.2	979	460	107	84.1
2RP2	1552 Concession 2	Tank	7.13	1.64	18.9	463	258	87.7	80.0
2RP3	1552 Concession 2	Effluent	7.55	1.27	18.6	191	28	82.6	76.0
3RP1	28 Highway 8	Inlet	7.53	2.25	14.2	173	46	14.6	12.3
3RP2*	28 Highway 8	Tank	7.59	0.65	15.4	208	24	24.4	19.6
3RP3	28 Highway 8	Effluent	7.67	2.93	15.0	199	33	24.0	18.9
4RP1	2649 No. 2 Sideroad	Inlet	7.18	1.82	11.8	219	202	30.6	24.1

Sample Code[1]	Location	Sample Type	pH	DO (mg/L)	Temp. (°C)	COD (mg/L)	TSS (mg/L)	TKN (mg/L)	NH3 (mg/L)
4RP2*	2649 No. 2 Sideroad	Tank	7.35	1.68	12.6	174	17	42.7	39.3
4RP3	2649 No. 2 Sideroad	Effluent	7.58	4.57	12.4	154	18	41.2	36.5
5RP1	4483 Escarpment Dr.	Inlet	6.31	0.29	14.1	14070	7480	266	131
5RP2*	4483 Escarpment Dr.	Tank	6.73	0.29	13.5	1133	705	78.7	59.1
5RP3	4483 Escarpment Dr.	Effluent	7.03	0.80	12.8	396	79	55.8	45.5
6RP1	8465 Canyon RI	Inlet	7.62	0.78	13.4	935	431	177	146
6RP2	8465 Canyon RI	Tank	7.45	1.18	13.5	696	60	141	122
6RP3	8465 Canyon RI	Effluent	7.56	4.21	13.3	682	52	139	122
1SP1	10091 Iona RD.	Inlet	7.69	1.19	22.7	1110	248	24.9	13.5
1SP2*	10091 Iona RD.	Tank	7.58	0.37	19.5	263	34	47.6	42.6
1SP3	10091 Iona RD.	Effluent	7.59	0.35	19.3	259	46	50.6	45.7
3SP1	1128 Matthiasville RD.	Inlet	6.51	5.10	11.5	21080	5470	399	81.1
3SP2	1128 Matthiasville RD.	Tank	7.76	0.79	8.9	718	137	64.8	48.9
3SP3	1128 Matthiasville RD.	Effluent	7.29	1.36	9.7	424	21	50.1	48.5
4SP1	493 Roaslind Lake Road	Inlet	9.40	4.28	7.8	580	30	228	207
4SP2*	493 Roaslind Lake Road	Tank	8.86	1.71	4.2	279	24	85.5	77.2
4SP3	493 Roaslind Lake Road	Effluent	8.82	3.51	4.3	263	15	86.8	78.2
5SP1	1002 Golden Point Road	Inlet	5.61	1.10	8.7	11890	4020	331	68.8
5SP2	1002 Golden Point Road	Tank	6.98	1.94	8.7	116	8	54.7	48.7
5SP3	1002 Golden Point Road	Effluent	6.95	2.35	8.9	109	5	55	49.2

Sample Code[1]	Location	Sample Type	pH	DO (mg/L)	Temp. (ºC)	COD (mg/L)	TSS (mg/L)	TKN (mg/L)	NH3 (mg/L)
6SP1	1206 Charlie Thompson RD.	Inlet	6.37	1.10	9.1	11310	3800	360	92.2
6SP2*	1206 Charlie Thompson RD.	Tank	6.98	1.94	8.9	571	74	155	132
6SP3	1206 Charlie Thompson RD.	Effluent	6.95	2.35	8.9	507	44	139	120
1RC1	44 Autumn Circle	Influent	7.22	1.76	11.6	514	70	75.7	65.6
1RC2	44 Autumn Circle	Tank	7.21	1.34	11.6	489	33	75.4	71.4
1RC3	44 Autumn Circle	Effluent	8.25	0.57	10.8	264	39	73.0	62.6
2RC1	7 Diamondwood Drive	Influent	6.72	0.61	16.9	950	650	98.9	86.1
2RC2	7 Diamondwood Drive	Tank	7.03	0.75	16.0	308	44	82.8	79.4
2RC3	7 Diamondwood Drive	Effluent	7.38	1.25	15.5	255	54	74.2	68.4
3RC1	10 Diamondwood Drive	Influent	7.09	0.65	17.7	739	424	75.8	50.2
3RC2	10 Diamondwood Drive	Tank	6.99	0.73	17.1	308	150	62.0	59.1
3RC3	10 Diamondwood Drive	Effluent	7.16	1.95	16.6	113	21	54.1	51.7
4RC1	17 Flamborough Hills Drive	Influent	7.37	1.23	13.2	615	48	67.9	59.5
4RC2*	17 Flamborough Hills Drive	Tank	7.42	1.62	12.7	504	83	70.8	60.3
4RC3	17 Flamborough Hills Drive	Effluent	7.39	3.20	13.2	431	66	71.6	60.1
5RC1	40 Diamondwood Drive	Influent	7.26	0.65	17.0	342	124	44.8	37.6
5RC2*	40 Diamondwood Drive	Tank	7.12	0.70	16.7	216	79	42.5	38.5
5RC3	40 Diamondwood Drive	Effluent	7.38	1.29	16.5	173	56	42.3	37.9
6RC1	45 Autumn Circle	Influent	7.11	1.63	11.5	380	52	78.5	67.7
6RC2*	45 Autumn Circle	Tank	7.09	1.41	11.1	323	48	81.9	72.0

Sample Code[1]	Location	Sample Type	pH	DO (mg/L)	Temp. (ºC)	COD (mg/L)	TSS (mg/L)	TKN (mg/L)	NH3 (mg/L)
6RC3	45 Autumn Circle	Effluent	7.21	0.42	10.5	271	44	79.8	67.8
1SC1	2 Diamondwood Drive	Influent	7.22	1.76	11.6	739	282	190	64.8
1SC2*	2 Diamondwood Drive	Tank	7.21	1.34	11.6	5864	2960	64.4	57.9
1SC3	2 Diamondwood Drive	Effluent	8.25	0.57	10.8	193	35	61.4	56.9
2SC1	22 Diamondwood Drive	Influent	6.72	0.61	16.9	505	163	52.7	44.0
2SC2*	22 Diamondwood Drive	Tank	7.03	0.75	16.0	381	125	44.9	37.9
2SC3	22 Diamondwood Drive	Effluent	7.38	1.25	15.5	255	50	39.5	34.8
3SC1	925 Longfellow Ave.	Influent	7.09	0.65	17.7	256	37	68.1	62.1
3SC2	925 Longfellow Ave.	Tank	6.99	0.73	17.1	211	42	70.7	67.2
3SC3	925 Longfellow Ave.	Effluent	7.16	1.95	16.6	127	26	64.4	61.2
4SC1	362 Evert Street	Influent	7.37	1.23	13.2	558	131	51.7	41.2
4SC2*	362 Evert Street	Tank	7.42	1.62	12.7	394	68	52.5	43.2
4SC3	362 Evert Street	Effluent	7.39	3.20	13.2	435	153	53.3	45.4
5SC1	3105 Dundas St.	Influent	7.26	0.65	17.0	441	432	68.4	55.8
5SC2	3105 Dundas St.	Tank	7.12	0.70	16.7	255	27	69.9	49.9
5SC3	3105 Dundas St.	Effluent	7.38	1.29	16.5	246	19	60.8	56.3
6SC1	2850 Victoria Street	Influent	7.11	1.63	11.5	486	200	47.1	35.0
6SC2	2850 Victoria Street	Tank	7.09	1.41	11.1	218	35	39.1	30.5
6SC3	2850 Victoria Street	Effluent	7.21	0.42	10.5	375	200	47.1	32.2

1. System types are RC (recirculating conventional), SC (single-pass conventional), RP (recirculating plug flow), and SP (single-pass plug flow).

* Samples included in comparing household onsite wastewater treatment system metagenomic analysis using 16S and shotgun sequencing.

Table 5. Total taxa identified at multiple stages of data treatment by 16S and shotgun sequencing.

Sequencing Method	Raw Identifications	Genera	Above 0.05% Relative Abundance	Genera Above 0.05% Relative Abundance	Above 0.1% Relative Abundance	Genera Above 0.1% Relative Abundance	Above 0.2% Relative Abundance	Genera Above 0.2% Relative Abundance
16S	3365	809	253	185	149	123	86	79
Shotgun	23819	3045	226	192	121	113	72	71

Table 6. Average taxa identified per sample at multiple stages of data treatment by 16S and shotgun sequencing.

Sequencing Method	Raw Identifications	Predicted True Species Count	Genera	Above 0.05% Relative Abundance	Genera Above 0.05% Relative Abundance	Above 0.1% Relative Abundance	Genera Above 0.1% Relative Abundance	Above 0.2% Relative Abundance	Genera Above 0.2% Relative Abundance
16S	550	558	277	131	136	88	96	56	65
Shotgun	12837	15738	2505	223	191	120	113	71	71

Table 7. Shotgun Significantly Differentially Abundant Taxa: Plug Flow Compared to Conventional.

Taxa	Base Mean	log2 Fold Change	Log2 Fold Change Standard Error	Stat	P-Value	Adjusted P-Value
Desulfomicrobium	6510	-3.85	0.836	-4.6	4.14E-06	3.52E-04
Azospirillum	1250	2.95	0.789	3.73	1.88E-04	0.00798

Table 8. 16S Significantly Differentially Abundant Taxa: Plug Flow Compared to Conventional.

Taxa	Base Mean	Log2 Fold Change	Log2 Fold Change Standard Error	Stat	P-Value	Adjusted P-Value
Subgroup_7	532	-9.52	2.01	-4.74	2.18E-06	2.11E-04
Desulfomicrobium	1590	-4.58	1.16	-3.95	7.74E-05	0.00375
Syner-01	368	-8.93	2.35	-3.79	1.49E-04	0.00482
Anaeroarcus	283	4.83	1.39	3.47	5.26E-04	0.0128
Dechloromonas	371	-4.95	1.63	-3.04	0.00238	0.0461

Table 9. Shotgun Significantly Differentially Abundant Taxa: Single-Pass Compared to Recirculating.

Taxa	Base Mean	Log2 Fold Change	Log2 Fold Change Standard Error	Stat	P-Value	Adjusted P-Value
Pusillimonas	14900	7.47	1.21	6.15	7.74E-10	6.58E-08
Phenylobacterium	9490	4.83	1.05	4.61	4.07E-06	1.15E-04
Simplicispira	21200	5.32	1.14	4.66	3.23E-06	1.15E-04
Geobacter	17000	-5.06	1.13	-4.48	7.56E-06	1.61E-04
Candidatus Cloacimonetes	1250	-3.54	0.979	-3.61	3.04E-04	0.00517
Rhizobiales	2240	3.20	0.902	3.55	3.86E-04	0.00548
Ruminococcus	4790	-1.80	0.618	-2.91	0.0036	0.0438

Table 10. 16S Significantly Differentially Abundant Taxa: Single-Pass Compared to Recirculating.

Taxa	Base Mean	Log2 Fold Change	Log2 Fold Change Standard Error	Stat	P-Value	Adjusted P-Value
Subgroup_7	532	26.9	2.03	13.3	2.61E-40	2.53E-38
Camelimonas	396	12.3	2.05	5.99	2.09E-09	1.01E-07
Hydrogenedensaceae	171	5.63	1.25	4.51	6.55E-06	2.12E-04
Burkholderiaceae	1760	3.92	0.889	4.41	1.04E-05	2.47E-04
Phenylobacterium	328	5.97	1.37	4.36	1.27E-05	2.47E-04

Taxa	Base Mean	Log2 Fold Change	Log2 Fold Change Standard Error	Stat	P-Value	Adjusted P-Value
Simplicispira	516	5.96	1.46	4.09	4.31E-05	6.97E-04
Brevundimonas	247	4.49	1.29	3.47	5.14E-04	0.00713
p-251-o5	115	-5.23	1.62	-3.23	0.00123	0.0149
Pleomorphomonas	202	4.73	1.55	3.05	0.0023	0.0248
Comamonas	244	3.88	1.31	2.95	0.00317	0.0307
Ruminococcus_2	146	-1.67	0.591	-2.83	0.00462	0.0356
Bact-08	320	-3.85	1.36	-2.82	0.00477	0.0356
Shinella	304	5.17	1.83	2.83	0.00468	0.0356

22

Literature Cited.

Adeyemi, F.M., Akinde, S.B., Oyedara, O.O., Wahab, A.A., Jimoh, S.O., 2019. Investigation of Escherichia coli Distribution in Drinking Water Wells Close to Septic Tanks in Densely Populated Areas of Osogbo, Osun State, Nigeria. Nigerian Journal of Microbiology 33, 4403–4414.

Afiahayati, Sato, K., Sakakibara, Y., 2015. MetaVelvet-SL: An Extension of the Velvet Assembler to a De Novo Metagenomic Assembler Utilizing Supervised Learning. DNA Research 22, 69–77. https://doi.org/10.1093/dnares/dsu041

Al-Yamani, F.Y., Polikarpov, I., Saburova, M., 2020. Marine Life Mortalities and Harmful Algal Blooms in the Northern Arabian Gulf. Aquatic Ecosystem Health and Management 23, 196–209. https://doi.org/10.1080/14634988.2020.1798157

Almeida, O.G.G., De Martinis, E.C.P., 2018. Bioinformatics Tools to Assess Metagenomic Data for Applied Microbiology. Applied Microbiology and Biotechnology 103, 69–82. https://doi.org/10.1007/s00253-018-9464-9

Amir, A., Daniel, M., Navas-Molina, J., Kopylova, E., Morton, J., Xu, Z.Z., Eric, K., Thompson, L., Hyde, E., Gonzalez, A., Knight, R., 2017. Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. American Society for Microbiology 2, 1–7.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., Brodie, E.L., Williams, K.H., Hubbard, S.S., Banfield, J.F., 2016. Thousands of Microbial Genomes Shed Light on Interconnected Biogeochemical Processes in an Aquifer System. Nature Communications 7, 1–11. https://doi.org/10.1038/ncomms13219

Anderson, M.J., 2017. Permutational Multivariate Analysis of Variance (PERMANOVA). Wiley

StatsRef: Statistics Reference Online 1–15. https://doi.org/10.1002/9781118445112.stat07841

Andrews, S., Krueger, F., Seconds-Pichon, A., Biggins, F., Wingett, S., 2015. FASTQC A Quality Control tool for High Throughput Sequence Data [WWW Document]. Babraham Institute. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S., Suzuki, M.T., 2017. Multiple Streptomyces Species with Distinct Secondary Metabolomes have Identical 16S rRNA Gene Sequences. Scientific Reports 7, 1–8. https://doi.org/10.1038/s41598-017-11363-1

Anukam, A., Mohammadi, A., Naqvi, M., Granström, K., 2019. A Review of the Chemistry of Anaerobic Digestion: Methods of Accelerating and Optimizing Process Efficiency. Processes 7, 1–19. https://doi.org/10.3390/PR7080504

Appling, D., Habteselassie, M.Y., Radcliffe, D., Bradshaw, J.K., 2013. Preliminary Study on the Effect of Wastewater Storage in Septic Tank on E. coli Concentration in Summer. Water (Switzerland) 5, 1141–1151. https://doi.org/10.3390/w5031141

Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4fun: Predicting Functional Profiles From Metagenomic 16S rRNA Data. Bioinformatics 31, 2882–2884. https://doi.org/10.1093/bioinformatics/btv287

Ayling, M., Clark, M.D., Leggett, R.M., 2019. New Approaches for Metagenome Assembly with Short Reads. Briefings in Bioinformatics 00, 1–11. https://doi.org/10.1093/bib/bbz020

Bag, S., Saha, B., Mehta, O., Anbumani, D., Kumar, N., Dayal, M., Pant, A., Kumar, P., Saxena, S., Allin, K.H., Hansen, T., Arumugam, M., Vestergaard, H., Pedersen, O., Pereira, V., Abraham, P., Tripathi, R., Wadhwa, N., Bhatnagar, S., Prakash, V.G., Radha, V., Anjana, R.M., Mohan, V., Takeda, K., Kurakawa, T., Nair, G.B., Das, B., 2016. An Improved Method

for High Quality Metagenomics Dna Extraction From Human and Environmental Samples. Scientific Reports 6, 1–9. https://doi.org/10.1038/srep26775

Batstone, D.J., 2006. Mathematical Modelling of Anaerobic Reactors Treating Domestic Wastewater: Rational Criteria for Model Use. Reviews in Environmental Science and Biotechnology 5, 57–71. https://doi.org/10.1007/s11157-005-7191-z

Batstone, D.J., Keller, J., Angelidaki, I., Kalyuzhnyi, S. V., Pavlostathis, S.G., Rozzi, A., Sanders, W.T., Siegrist, H., Vavilin, V.A., 2002. The IWA Anaerobic Digestion Model No 1 (ADM1). Water science and technology : a journal of the International Association on Water Pollution Research 45, 65–73. https://doi.org/10.2166/wst.2002.0292

Batstone, D.J., Puyol, D., Flores-Alsina, X., Rodríguez, J., 2015. Mathematical Modelling of Anaerobic Digestion Processes: Applications and Future Needs. Reviews in Environmental Science and Biotechnology 14, 595–613. https://doi.org/10.1007/s11157-015-9376-4

Bertrand, D., Shaw, J., Kalathiyappan, M., Ng, A.H.Q., Kumar, M.S., Li, C., Dvornicic, M., Soldo, J.P., Koh, J.Y., Tong, C., Ng, O.T., Barkham, T., Young, B., Marimuthu, K., Chng, K.R., Sikic, M., Nagarajan, N., 2019. Hybrid Metagenomic Assembly Enables High-Resolution Analysis of Resistance Determinants and Mobile Elements in Human Microbiomes. Nature Biotechnology 37, 937–944. https://doi.org/10.1038/s41587-019-0191-2

Binod, P., Sindhu, R., Pandey, A., 2013. The Alcohol Fermentation Step: The Most Common Ethanologenic Microorganisms Among Yeasts, Bacteria and Filamentous Fungi. Lignocellulose Conversion: Enzymatic and Microbial Tools for Bioethanol Production 131–149. https://doi.org/10.1007/978-3-642-37861-4_7

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. Bioinformatics 30, 2114–2120.

https://doi.org/10.1093/bioinformatics/btu170

Boyd, E.S., Amenabar, M.J., Poudel, S., Templeton, A.S., 2020. Bioenergetic Constraints on the Origin of Autotrophic Metabolism. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. https://doi.org/10.1098/rsta.2019.0151

Bray, J.R., Curtis, J.T., 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Society of America 27, 325–349.

Breitwieser, F.P., Lu, J., Salzberg, S.L., 2018. A Review of Methods and Databases for Metagenomic Classification and Assembly. Briefings in Bioinformatics 20, 1125–1139. https://doi.org/10.1093/bib/bbx120

Brown, C.M., Staley, C., Wang, P., Dalzell, B., Chun, C.L., Sadowsky, M.J., 2017. A High-Throughput DNA-Sequencing Approach for Determining Sources of Fecal Bacteria in a Lake Superior Estuary. Environmental Science and Technology 51, 8263–8271. https://doi.org/10.1021/acs.est.7b01353

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., Banfield, J.F., 2015. Unusual Biology Across a Group Comprising More Than 15% Of Domain Bacteria. Nature 523, 208–211. https://doi.org/10.1038/nature14486

Buchfink, B., Xie, C., Huson, D.H., 2014. Fast and Sensitive Protein Alignment Using Diamond. Nature Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176

Bushnell, B., 2020. BBMap [WWW Document]. Source Forge. URL https://sourceforge.net/projects/bbmap/ (accessed 4.22.20).

Bushnell, B., Rood, J., Singer, E., 2017. BBMerge – Accurate Paired Shotgun Read Merging via Overlap. PLoS ONE 12, 1–15. https://doi.org/10.1371/journal.pone.0185056

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis. ISME Journal 11, 2639–2643. https://doi.org/10.1038/ismej.2017.119

Callahan, Benjamin J, Mcmurdie, P.J., Rosen, M.J., Han, A.W., A, A.J., 2016. DADA2: High Resolution Sample Inference From Illumina Amplicon Data. Nature Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869.DADA2

Callahan, Benjamin J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: High-Resolution Sample Inference From Illumina Amplicon Data. Nature Methods 13, 581–583. https://doi.org/10.1038/nmeth.3869

Caminade, C., McIntyre, K.M., Jones, A.E., 2019. Impact of Recent and Future Climate Change on Vector-Borne Diseases. Annals of the New York Academy of Sciences 1436, 157–173. https://doi.org/10.1111/nyas.13950

Capson-Tojo, G., Moscoviz, R., Astals, S., Robles, Steyer, J.P., 2020. Unraveling the Literature Chaos Around Free Ammonia Inhibition in Anaerobic Digestion. Renewable and Sustainable Energy Reviews. https://doi.org/10.1016/j.rser.2019.109487

Capson-Tojo, G., Moscoviz, R., Ruiz, D., Santa-Catalina, G., Trably, E., Rouez, M., Crest, M., Steyer, J.P., Bernet, N., Delgenès, J.P., Escudié, R., 2018a. Addition of Granular Activated Carbon and Trace Elements to Favor Volatile Fatty Acid Consumption During Anaerobic Digestion of Food Waste. Bioresource Technology 260, 157–168. https://doi.org/10.1016/j.biortech.2018.03.097

Capson-Tojo, G., Trably, E., Rouez, M., Crest, M., Bernet, N., Steyer, J.P., Delgenès, J.P., Escudié, R., 2018b. Methanosarcina Plays a Main Role During Methanogenesis of High-Solids Food Waste and Cardboard. Waste Management 76, 423–430.

https://doi.org/10.1016/j.wasman.2018.04.004

Chakravorty, S., Helb, D., Burday, M., Connell, N., Alland, D., 2007. A Detailed Analysis of 16S Ribosomal RNA Gene Segments for the Diagnosis of Pathogenic Bacteria. Journal of Microbiological Methods 69, 330–339. https://doi.org/10.1016/j.mimet.2007.02.005

Chamberlain, J.S., Gibbs, R.A., Rainer, J.E., Nguyen, P.N., Thomas, C., 1988. Deletion Screening of the Duchenne Muscular Dystrophy Locus via Multiplex Dna Amplification. Nucleic Acids Research 16, 11141–11156. https://doi.org/10.1093/nar/16.23.11141

Chan, A.W.Y., Naphtali, J., Schellhorn, H.E., 2019. High-Throughput Dna Sequencing Technologies for Water and Wastewater Analysis. Science Progress. https://doi.org/10.1177/0036850419881855

Chan, W.Y.A., 2020. An Investigation of Potential Relationships Between Septic Tank Microbial Communities and System Design and Performance. McMaster University.

Chen, H., Wang, Z., Liu, H., Nie, Y., Zhu, Y., Jia, Q., Ding, G., Ye, J., 2021. Variable Sediment Methane Production in Response to Different Source-Associated Sewer Sediment Types and Hydrological Patterns: Role of the Sediment Microbiome. Water Research 190, 116670. https://doi.org/10.1016/j.watres.2020.116670

Clooney, A.G., Fouhy, F., Sleator, R.D., O'Driscoll, A., Stanton, C., Cotter, P.D., Claesson, M.J., 2016. Comparing Apples and Oranges?: Next Generation Sequencing and its Impact on Microbiome Analysis. PLoS ONE 11, 1–16. https://doi.org/10.1371/journal.pone.0148028

Connelly, S., Pussayanavin, T., Randle-Boggis, R.J., Wicheansan, A., Jampathong, S., Keating, C., Ijaz, U.Z., Sloan, W.T., Koottatep, T., 2019. Solar Septic Tank: Next Generation Sequencing Reveals Effluent Microbial Community Composition as a Useful Index of System Performance. Water (Switzerland) 11, 1–14. https://doi.org/10.3390/W11122660

Connelly, S., Shin, S.G., Dillon, R.J., Ijaz, U.Z., Quince, C., Sloan, W.T., Collins, G., 2017. Bioreactor Scalability: Laboratory-Scale Bioreactor Design Influences Performance, Ecology, and Community Physiology in Expanded Granular Sludge Bed Bioreactors. Frontiers in Microbiology 8. https://doi.org/10.3389/fmicb.2017.00664

Cordero, O.X., Datta, M.S., 2016. Microbial Interactions and Community Assembly at Microscales. Current Opinion in Microbiology 31, 227–234. https://doi.org/10.1016/j.mib.2016.03.015

De Clercq, D., Wen, Z., Fei, F., Caicedo, L., Yuan, K., Shang, R., 2020. Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion. Science of the Total Environment 712, 134574. https://doi.org/10.1016/j.scitotenv.2019.134574

de Jonge, N., Davidsson, Å., la Cour Jansen, J., Nielsen, J.L., 2020. Characterisation of Microbial Communities for Improved Management of Anaerobic Digestion of Food Waste. Waste Management 117, 124–135. https://doi.org/10.1016/j.wasman.2020.07.047

DeFlorio-Barker, S., Wing, C., Jones, R.M., Dorevitch, S., 2018. Estimate of Incidence and Cost of Recreational Waterborne Illness on United States Surface Waters. Environmental Health: A Global Access Science Source 17, 1–10. https://doi.org/10.1186/s12940-017-0347-9

Dixon, P., 2003. Vegan, a Package of R Functions for Community Ecology. Journal of Vegetation Science 14, 927–930.

Duan, Y., Shi, F., 2014. Bioreactor Design for Algal Growth as a Sustainable Energy Source, in: Reactor and Process Design in Sustainable Energy Technology. Elsevier B.V., pp. 27–60. https://doi.org/10.1016/B978-0-444-59566-9.00002-8

Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., De Cesare, A., 2021.

Comparison Between 16s Rrna and Shotgun Sequencing Data for the Taxonomic Characterization of the Gut Microbiota. Scientific Reports | 11, 3030. https://doi.org/10.1038/s41598-021-82726-y

Edgar, R.C., 2018. Updating the 97% Identity Threshold for 16S Ribosomal RNA OTUs. Bioinformatics 34, 2371–2375. https://doi.org/10.1093/bioinformatics/bty113

Edgar, Robert C., 2016. UCHIME2: Improved Chimera Prediction for Amplicon Sequencing. bioRxiv 074252. https://doi.org/10.1101/074252

Edgar, Robert C, 2016. UNOISE2: Improved Error-Correction for Illumina 16S and ITS Amplicon Sequencing. bioRxiv 081257. https://doi.org/10.1101/081257

Edgar, R.C., 2013. Uparse: Highly Accurate Otu Sequences From Microbial Amplicon Reads. Nature Methods 10, 996–998. https://doi.org/10.1038/nmeth.2604

Escobar-Zepeda, A., Godoy-Lozano, E.E., Raggi, L., Segovia, L., Merino, E., Gutiérrez-Rios, R.M., Juarez, K., Licea-Navarro, A.F., Pardo-Lopez, L., Sanchez-Flores, A., 2018. Analysis of Sequencing Strategies and Tools for Taxonomic Annotation: Defining Standards for Progressive Metagenomics. Scientific Reports 8, 1–13. https://doi.org/10.1038/s41598-018-30515-5

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: Summarize Analysis Results For Multiple Tools and Samples in a Single Report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fu, B., Jin, X., Conrad, R., Liu, Hongbo, Liu, He, 2019. Competition Between Chemolithotrophic Acetogenesis and Hydrogenotrophic Methanogenesis for Exogenous h2 /CO2 in Anaerobically Digested Sludge: Impact of Temperature. Frontiers in Microbiology. https://doi.org/10.3389/fmicb.2019.02418

Gajer, P., He, X., Mccomb, E., Gravitt, P.E., Ghanem, K.G., Brotman, R.M., 2021. Ultrahigh-Throughput Multiplexing and Sequencing of >500- Base-Pair Amplicon Regions on the Illumina HiSeq 2500 Platform. Applied and Environmental Science 4.

García-Lozano, M., Hernández-De Lira, I.O., Huber, D.H., Balagurusamy, N., 2019. Spatial variations of bacterial communities of an anaerobic lagoon-type biodigester fed with dairy manure. Processes 7. https://doi.org/10.3390/pr7070408

Garcia, S.L., Buck, M., Hamilton, J.J., Wurzbacher, C., Grossart, H.-P., McMahon, K.D., Eiler, A., 2018. Model Communities Hint at Promiscuous Metabolic Linkages between Ubiquitous Free-Living Freshwater Bacteria. mSphere 3, 1–8. https://doi.org/10.1128/msphere.00202-18

Gaur, M., Vasudeva, A., Singh, A., Sharma, V., Khurana, H., Negi, R.K., Lee, J.K., Kalia, V.C., Misra, R., Singh, Y., 2019. Comparison of DNA Extraction Methods for Optimal Recovery of Metagenomic DNA from Human and Environmental Samples. Indian Journal of Microbiology 59, 482–489. https://doi.org/10.1007/s12088-019-00832-y

Gonze, D., Coyte, K.Z., Lahti, L., Faust, K., 2018. Microbial Communities as Dynamical Systems. Current Opinion in Microbiology 44, 41–49. https://doi.org/10.1016/j.mib.2018.07.004

Grattan, L.M., Holobaugh, S., Morris, J.G., 2016. Harmful Algal Blooms and Public Health. Harmful Algae 57, 2–8. https://doi.org/10.1016/j.hal.2016.05.003

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D. V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methé, B., DeSantis, T.Z., Petrosino, J.F., Knight, R., Birren, B.W., 2011. Chimeric 16s rRNA Sequence Formation and Detection in Sanger and 454-Pyrosequenced PCR Amplicons. Genome Research 21, 494–504. https://doi.org/10.1101/gr.112730.110

Hassan, M., Essam, T., Mira, A., Megahed, S., 2019. Biomonitoring Detoxification Efficiency of

an Algal-Bacterial Microcosm System for Treatment of Coking Wastewater: Harmonization Between Chlorella Vulgaris Microalgae and Wastewater Microbiome. Science of the Total Environment 677, 120–130. https://doi.org/10.1016/j.scitotenv.2019.04.304

He, Q., Li, L., Peng, X., 2017. Early Warning Indicators and Microbial Mechanisms for Process Failure due to Organic Overloading in Food Waste Digesters. Journal of Environmental Engineering 143, 04017077. https://doi.org/10.1061/(asce)ee.1943-7870.0001280

Herbold, C.W., Pelikan, C., Kuzyk, O., Hausmann, B., Angel, R., Berry, D., Loy, A., 2015. A Flexible and Economical Barcoding Approach for Highly Multiplexed Amplicon Sequencing of Diverse Target Genes. Frontiers in Microbiology 6, 1–8. https://doi.org/10.3389/fmicb.2015.00731

Herrero, M., Stuckey, D.C., 2014. Bioaugmentation and Its Application in Wastewater Treatment: A Review. Chemosphere 140, 119–128. https://doi.org/10.1016/j.chemosphere.2014.10.033

Hidalgo, K.J., Saito, T., Silva, R.S., Delforno, T.P., Duarte, I.C.S., de Oliveira, V.M., Okada, D.Y., 2021. Microbiome Taxonomic and Functional Profiles of Two Domestic Sewage Treatment Systems. Biodegradation 32, 17–36. https://doi.org/10.1007/s10532-020-09921-y

Hsieh, T.C., Ma, K.H., Chao, A., 2016. iNEXT: an R Package for Rarefaction and Extrapolation of Species Diversity (Hill numbers). Methods in Ecology and Evolution. https://doi.org/10.1111/2041-210X.12613

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., Von Mering, C., Bork, P., 2019. Eggnog 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses. Nucleic Acids Research 47, D309–D314. https://doi.org/10.1093/nar/gky1085

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., Ruscheweyh, H.J., Tappu, R., 2016. Megan Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Computational Biology 12, 1–12. https://doi.org/10.1371/journal.pcbi.1004957

Hyatt, D., Chen, G.L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. BMC Bioinformatics 11. https://doi.org/10.1186/1471-2105-11-119

Ibrahim, S., Azab El-Liethy, M., Abia, A.L.K., Abdel-Gabbar, M., Mahmoud Al Zanaty, A., Mohamed Kamel, M., 2020. Design of a Bioaugmented Multistage Biofilter for Accelerated Municipal Wastewater Treatment and Deactivation of Pathogenic Microorganisms. Science of the Total Environment 703, 134786. https://doi.org/10.1016/j.scitotenv.2019.134786

Jakaitiene, A., Sangiovanni, M., Guarracino, M.R., Pardalos, P.M., 2016. Multidimensional Scaling for Genomic Data, in: Advances in Stochastic and Deterministic Global Optimization. Springer International Publishing, pp. 229–242. https://doi.org/10.1007/978-3-319-29975-4

Jansson, J.K., Hofmockel, K.S., 2018. The Soil Microbiome — From Metagenomics to Metaphenomics. Current Opinion in Microbiology 43, 162–168. https://doi.org/10.1016/j.mib.2018.01.013

Jiang, X.T., Ye, L., Ju, F., Li, B., Ma, L.P., Zhang, T., 2018. Temporal Dynamics of Activated Sludge Bacterial Communities in Two Diversity Variant Full-Scale Sewage Treatment Plants. Applied Microbiology and Biotechnology 102, 9379–9388. https://doi.org/10.1007/s00253-018-9287-8

Jiang, Y., Banks, C., Zhang, Y., Heaven, S., Longhurst, P., 2017. Quantifying the Percentage of

Methane Formation via Acetoclastic and Syntrophic Acetate Oxidation Pathways in Anaerobic Digesters. Waste Management 71, 749–756. https://doi.org/10.1016/j.wasman.2017.04.005

Johnson, J., Jain, K., Madamwar, D., 2017. Functional Metagenomics: Exploring Nature's Gold Mine. Current Developments in Biotechnology and Bioengineering: Functional Genomics and Metabolic Engineering 27–43. https://doi.org/10.1016/B978-0-444-63667-6.00002-X

Johnson, J.S., Spakowicz, D.J., Hong, B.Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M., 2019. Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis. Nature Communications 10, 1–11. https://doi.org/10.1038/s41467-019-13036-1

Jourdin, L., Freguia, S., Donose, B.C., Keller, J., 2015. Autotrophic Hydrogen-Producing Biofilm Growth Sustained by a Cathode as the Sole Electron and Energy Source. Bioelectrochemistry 102, 56–63. https://doi.org/10.1016/j.bioelechem.2014.12.001

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., Wong, G.K.S., 2016. Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. Frontiers in Microbiology 7, 1–17. https://doi.org/10.3389/fmicb.2016.00459

Jowett, C., Kraemer, J., James, C., Jowett, E.C., 2017. Improving Septic Tank Performance By Enhancing Anaerobic Digestion, in: National Onsite Wastewater Recycling Association Proceedings.

Kanehisa, M., 2004. The Kegg Resource for Deciphering the Genome. Nucleic Acids Research 32, 277D – 280. https://doi.org/10.1093/nar/gkh063

Karekar, S.C., Srinivas, K., Ahring, B.K., 2019. Kinetic study on Heterotrophic Growth of

299 Acetobacterium Woodii on Lignocellulosic Substrates for Acetic Acid Production.
300 Fermentation 5. https://doi.org/10.3390/fermentation5010017
301 Karlson, B., Andersen, P., Arneborg, L., Cembella, A., Eikrem, W., John, U., West, J.J., Klemm,
302 K., Kobos, J., Lehtinen, S., Lundholm, N., Mazur-Marzec, H., Naustvoll, L., Poelman, M.,
303 Provoost, P., De Rijcke, M., Suikkanen, S., 2021. Harmful Algal Blooms and Their Effects
304 in Coastal Seas of Northern Europe. Harmful Algae 102, 101989.
305 https://doi.org/10.1016/j.hal.2021.101989
306 Kassambara, A., 2019. FASTQC: Quality Control of Sequencing Data (Version 0.1.2) [WWW
307 Document]. URL https://cran.r-project.org/web/packages/fastqcr/index.html (accessed
308 11.24.20).
309 Khachatryan, L., de Leeuw, R.H., Kraakman, M.E.M., Pappas, N., te Raa, M., Mei, H., de Knijff,
310 P., Laros, J.F.J., 2020. Taxonomic Classification and Abundance Estimation Using 16S and
311 WGS—A Comparison Using Controlled Reference Samples. Forensic Science International:
312 Genetics 46, 102257. https://doi.org/10.1016/j.fsigen.2020.102257
313 Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D., 2013. Development of
314 a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence
315 Data on the Miseq Illumina Sequencing Platform. Applied and Environmental Microbiology
316 79, 5112–5120. https://doi.org/10.1128/AEM.01043-13
317 Kultima, J.R., Coelho, L.P., Forslund, K., Huerta-Cepas, J., Li, S.S., Driessen, M., Voigt, A.Y.,
318 Zeller, G., Sunagawa, S., Bork, P., 2016. Mocat2: A Metagenomic Assembly, Annotation and
319 Profiling Framework. Bioinformatics 32, 2520–2523.
320 https://doi.org/10.1093/bioinformatics/btw183
321 Kurade, M.B., Saha, S., Salama, E.S., Patil, S.M., Govindwar, S.P., Jeon, B.H., 2019. Acetoclastic

Methanogenesis Led by Methanosarcina in Anaerobic Co-digestion of Fats, Oil and Grease for Enhanced Production of Methane. Bioresource Technology 272, 351–359. https://doi.org/10.1016/j.biortech.2018.10.047

Kurth, J.M., Op den Camp, H.J.M., Welte, C.U., 2020. Several Ways One Goal—Methanogenesis From Unconventional Substrates. Applied Microbiology and Biotechnology 104, 6839–6854. https://doi.org/10.1007/s00253-020-10724-7

Kushkevych, I. V., 2013. Identification of Sulfate-Reducing Bacteria Strains of Human Large Intestine. Studia Biologica 7, 115–132. https://doi.org/10.30970/sbi.0703.312

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., Huttenhower, C., 2013. Predictive Functional Profiling of Microbial Communities Using 16S rRNA Marker Gene Sequences. Nature Biotechnology 31, 814–821. https://doi.org/10.1038/nbt.2676

Lapointe, B.E., Herren, L.W., Paule, A.L., 2017. Septic Systems Contribute to Nutrient Pollution and Harmful Algal Blooms in the St. Lucie Estuary, Southeast Florida, USA. Harmful Algae 70, 1–22. https://doi.org/10.1016/j.hal.2017.09.005

Leigh Greathouse, K., Sinha, R., Vogtmann, E., 2019. DNA Extraction for Human Microbiome Studies: The Issue of Standardization. Genome Biology 20, 1–4. https://doi.org/10.1186/s13059-019-1843-8

Li, D., Huang, Y., Leung, C.M., Luo, R., Ting, H.F., Lam, T.W., 2017. Megagta: A Sensitive and Accurate Metagenomic Gene-Targeted Assembler Using Iterative de Bruijn Graphs. BMC Bioinformatics 18. https://doi.org/10.1186/s12859-017-1825-3

Li, D., Liu, C.M., Luo, R., Sadakane, K., Lam, T.W., 2015. MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn

Graph. Bioinformatics 31, 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, Y., Chen, Y., Wu, J., 2019. Enhancement of Methane Production in Anaerobic Digestion Process: A Review. Applied Energy 240, 120–137. https://doi.org/10.1016/j.apenergy.2019.01.243

Li, Y., Sun, Y., Li, L., Yuan, Z., 2018. Acclimation of Acid-Tolerant Methanogenic Propionate-Utilizing Culture and Microbial Community Dissecting. Bioresource Technology. https://doi.org/10.1016/j.biortech.2017.11.034

Li, Z., Liu, B., Cui, H., Ding, J., Li, H., Xie, G., Ren, N., Xing, D., 2019. The Complete Genome Sequence of Ethanoligenens Harbinense Reveals the Metabolic Pathway of Acetate-Ethanol Fermentation: A Novel Understanding of the Principles of Anaerobic Biotechnology. Environment International 131, 105053. https://doi.org/10.1016/j.envint.2019.105053

Li, Z., Lou, Y., Ding, J., Liu, B.F., Xie, G.J., Ren, N.Q., Xing, D., 2020. Metabolic Regulation of Ethanol-Type Fermentation of Anaerobic Acidogenesis at Different PH Based on Transcriptome Analysis of Ethanoligenens Harbinense. Biotechnology for Biofuels 13, 1–16. https://doi.org/10.1186/s13068-020-01740-w

Linz, A.M., Crary, B.C., Shade, A., Owens, S., Gilbert, J.A., Knight, R., McMahon, K.D., 2017. Bacterial Community Composition and Dynamics Spanning Five Years in Freshwater Bog Lakes. mSphere 2. https://doi.org/10.1128/msphere.00169-17

Ljungdahl, L.G., 1986. The Autotrophic Pathway of Acetogenic Bacteria. Ann. Rev. Microbiology 40, 415–450.

Long, F., Wang, L., Cai, W., Lesnik, K., Liu, H., 2021. Predicting the Performance of Anaerobic Digestion Using Machine Learning Algorithms and Genomic Data. Water Research 199, 117182. https://doi.org/10.1016/j.watres.2021.117182

Love, M.I., Huber, W., Anders, S., 2014. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data With Deseq2. Genome Biology 15, 1–21. https://doi.org/10.1186/s13059-014-0550-8

Ma, J., Frear, C., Wang, Z.W., Yu, L., Zhao, Q., Li, X., Chen, S., 2013. A Simple Methodology for Rate-Limiting Step Determination for Anaerobic Digestion of Complex Substrates and Effect of Microbial Community Ratio. Bioresource Technology 134, 391–395. https://doi.org/10.1016/j.biortech.2013.02.014

Ma, S. jia, Ma, H. jun, Hu, H. dong, Ren, H. qiang, 2019. Effect of MixingIintensity on Hydrolysis and Acidification of Sewage Sludge in Two-Stage Anaerobic Digestion: Characteristics of Dissolved Organic Matter and the Key Microorganisms. Water Research 148, 359–367. https://doi.org/10.1016/j.watres.2018.10.058

Mahajan, R., Attri, S., Sharma, K., Singh, N., Sharma, D., Goel, G., 2018. Statistical Assessment of Dna Extraction Methodology for Culture-Independent Analysis of Microbial Community Associated With Diverse Environmental Samples. Molecular Biology Reports 45, 297–308. https://doi.org/10.1007/s11033-018-4162-3

Mallin, M.A., McIver, M.R., 2012. Pollutant Impacts to Cape Hatteras National Seashore From Urban Runoff and Septic Leachate. Marine Pollution Bulletin 64, 1356–1366. https://doi.org/10.1016/j.marpolbul.2012.04.025

Manaia, C.M., Rocha, J., Scaccia, N., Marano, R., Radu, E., Biancullo, F., Cerqueira, F., Fortunato, G., Iakovides, I.C., Zammit, I., Kampouris, I., Vaz-Moreira, I., Nunes, O.C., 2018. Antibiotic Resistance in Wastewater Treatment Plants: Tackling the Black Box. Environment International 115, 312–324. https://doi.org/10.1016/j.envint.2018.03.044

Markowitz, V.M., Chen, I.M.A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A.,

Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C., 2012. IMG: The Integrated Microbial Genomes Database and Comparative Analysis System. Nucleic Acids Research 40, 115–122. https://doi.org/10.1093/nar/gkr1044

Martin, M., 2011. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads. EMBnet 17, 10–12. https://doi.org/https://doi.org/10.14806/ej.17.1.200

Mayer, A., Weuster-Botz, D., 2017. Reaction engineering analysis of the autotrophic energy metabolism of Clostridium aceticum. FEMS Microbiology Letters 364, 1–8. https://doi.org/10.1093/femsle/fnx219

McCombie, W.R., McPherson, J.D., Mardis, E.R., 2019. Next-Generation Sequencing Technologies. Cold Spring Harbor Perspectives in Medicine 9. https://doi.org/10.1101/cshperspect.a036798

McMurdie, P.J., Holmes, S., 2012. Phyloseq: A Bioconductor Package for Handling and Analysis of High-Throughput Phylogenetic Sequence Data. Pacific Symposium on Biocomputing 235–246. https://doi.org/10.1142/9789814366496_0023

Milanese, A., Mende, D.R., Paoli, L., Salazar, G., Ruscheweyh, H.J., Cuenca, M., Hingamp, P., Alves, R., Costea, P.I., Coelho, L.P., Schmidt, T.S.B., Almeida, A., Mitchell, A.L., Finn, R.D., Huerta-Cepas, J., Bork, P., Zeller, G., Sunagawa, S., 2019. Microbial Abundance, Activity and Population Genomic Profiling With mOTUs2. Nature Communications 10. https://doi.org/10.1038/s41467-019-08844-4

Milner, M.G., Curtis, T.P., Davenport, R.J., 2008. Presence and Activity of Ammonia-Oxidising Bacteria Detected Amongst the Overall Bacterial Diversity Along a Physico-Chemical Gradient of a Nitrifying Wastewater Treatment Plant. Water Research 42, 2863–2872.

https://doi.org/10.1016/j.watres.2008.02.019

Mohiuddin, M.M., Botts, S.R., Paschos, A., Schellhorn, H.E., 2019. Temporal and Spatial Changes in Bacterial Diversity in Mixed Use Watersheds of the Great Lakes Region. Journal of Great Lakes Research 45, 109–118. https://doi.org/10.1016/j.jglr.2018.10.007

Mohiuddin, M.M., Salama, Y., Schellhorn, H.E., Golding, G.B., 2017. Shotgun Metagenomic Sequencing Reveals Freshwater Beach Sands as Reservoir of Bacterial Pathogens. Water Research 115, 360–369. https://doi.org/10.1016/j.watres.2017.02.057

Mysara, M., Saeys, Y., Leys, N., Raes, J., Monsieurs, P., 2015. CATCH, an Ensemble Classifier for Chimera Detection in 16S rRNA Sequencing Studies. Applied and Environmental Microbiology 81, 1573–1584. https://doi.org/10.1128/AEM.02896-14

Najafpour, G.D., 2015. Biochemical Engineering and Biotechnology: Second Edition, Biochemical Engineering and Biotechnology: Second Edition. Elsevier Inc.

Naphtali, J., 2020. Microbial Communities in Septic Tank Anaerobic Digesters and Their Interactions with Digester Design and Chemical Environment. McMaster University.

Ni, J., Li, X., He, Z., Xu, M., 2017. A Novel Method to Determine the Minimum Number of Sequences Required for Reliable Microbial Community Analysis. Journal of Microbiological Methods 139, 196–201. https://doi.org/10.1016/j.mimet.2017.06.006

Nidhra, S., 2012. Black Box and White Box Testing Techniques - A Literature Review. International Journal of Embedded Systems and Applications 2, 29–50. https://doi.org/10.5121/ijesa.2012.2204

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. MetaSPAdes: A New Versatile Metagenomic Assembler. Genome Research 27, 824–834. https://doi.org/10.1101/gr.213959.116

O'Donnell, A.G., Young, I.M., Rushton, S.P., Shirley, M.D., Crawford, J.W., 2007. Visualization, Modelling and Prediction in Soil Microbiology. Nature Reviews Microbiology 5, 689–699. https://doi.org/10.1038/nrmicro1714

Oh, Y.S., Roh, D.H., 2012. Phenylobacterium Muchangponense sp. nov., Isolated from Beach Soil, and Emended Description of the Genus Phenylobacterium. International Journal of Systematic and Evolutionary Microbiology 62, 977–983. https://doi.org/10.1099/ijs.0.028902-0

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The Seed and the Rapid Annotation of Microbial Genomes Using Subsystems Technology (Rast). Nucleic Acids Research 42, 206–214. https://doi.org/10.1093/nar/gkt1226

Palamuleni, L.G., 2002. Effect of Sanitation Facilities, Domestic Solid Waste Disposal and Hygiene Practices on Water Quality in Malawiõs Urban Poor Areas: A Case Study of South Lunzu Township in the City of Blantyre. Physics and Chemistry of the Earth 27, 845–850.

Parada, A.E., Needham, D.M., Fuhrman, J.A., 2016. Every Base Matters: Assessing Small Subunit rRNA Primers for Marine Microbiomes With Mock Communities, Time Series and Global Field Samples. Environmental Microbiology 18, 1403–1414. https://doi.org/10.1111/1462-2920.13023

Peces, M., Astals, S., Jensen, P.D., Clarke, W.P., 2018. Deterministic Mechanisms Define the Long-Term Anaerobic Digestion Microbiome and its Functionality Regardless of the Initial Microbial Community. Water Research 141, 366–376. https://doi.org/10.1016/j.watres.2018.05.028

Peng, Y., Leung, H.C.M., Yiu, S.M., Chin, F.Y.L., 2012. IDBA-UD: A de Novo Assembler for

Single-Cell and Metagenomic Sequencing Data With Highly Uneven Depth. Bioinformatics 28, 1420–1428. https://doi.org/10.1093/bioinformatics/bts174

Pevzner, P.A., Tang, H., Waterman, M.S., 2001. An Eulerian path approach to DNA fragment assembly. Proceedings of the National Academy of Sciences of the United States of America 98, 9748–9753. https://doi.org/10.1073/pnas.171285098

Prodan, A., Tremaroli, V., Brolin, H., Zwinderman, A.H., Nieuwdorp, M., Levin, E., 2020. Comparing Bioinformatic Pipelines for Microbial 16S rRNA Amplicon Sequencing. PLoS ONE 15, 1–19. https://doi.org/10.1371/journal.pone.0227434

Puentes-Téllez, P.E., Salles, J.F., 2020. Dynamics of Abundant and Rare Bacteria During Degradation of Lignocellulose from Sugarcane Biomass. Microbial Ecology 79, 312–325. https://doi.org/10.1007/s00248-019-01403-w

Qi, G., Meng, W., Zha, J., Zhang, S., Yu, S., Liu, J., Ren, L., 2019. A novel insight into the influence of thermal pretreatment temperature on the anaerobic digestion performance of floatable oil-recovered food waste: Intrinsic transformation of materials and microbial response. Bioresource Technology. https://doi.org/10.1016/j.biortech.2019.122021

Quince, C., Walker, A.W., Simpson, J.T., Loman, N.J., Segata, N., 2017. Shotgun Metagenomics, From Sampling to Analysis. Nature Biotechnology 35, 833–844. https://doi.org/10.1038/nbt.3935

Ranjan, R., Rani, A., Metwally, A., McGee, H.S., Perkins, D.L., 2016. Analysis of the Microbiome: Advantages of Whole Genome Shotgun Versus 16S Amplicon Sequencing. Biochemical and Biophysical Research Communications 469, 967–977. https://doi.org/10.1016/j.bbrc.2015.12.083

Raynaud, X., Nunan, N., 2014. Spatial Ecology of Bacteria at the Microscale in Soil. PLoS ONE

9. https://doi.org/10.1371/journal.pone.0087217

Reymond, P., Abdel Wahaab, R., Moussa, M.S., Lüthi, C., 2018. Scaling up small scale wastewater treatment systems in low- and middle-income countries: An analysis of challenges and ways forward through the case of Egypt. Utilities Policy 52, 13–21. https://doi.org/10.1016/j.jup.2018.03.008

Richards, S., Withers, P.J.A., Paterson, E., McRoberts, C.W., Stutter, M., 2017. Potential Tracers for Tracking Septic Tank Effluent Discharges in Watercourses. Environmental Pollution 228, 245–255. https://doi.org/10.1016/j.envpol.2017.05.044

Roopnarain, A., Mukhuba, M., Adeleke, R., Moeletsi, M., 2017. Biases During Dna Extraction Affect Bacterial and Archaeal Community Profile of Anaerobic Digestion Samples. 3 Biotech 7, 1–12. https://doi.org/10.1007/s13205-017-1009-x

Rosen, M.J., Callahan, B.J., Fisher, D.S., Holmes, S.P., 2012. Denoising PCR-Amplified Metagenome Data. BMC Bioinformatics 13. https://doi.org/10.1186/1471-2105-13-283

Saetre, P., Bååth, E., 2000. Spatial Variation and Patterns of Soil Microbial Community Structure in a Mixed Spruce-Birch Stand. Soil Biology and Biochemistry 32, 909–917. https://doi.org/10.1016/S0038-0717(99)00215-1

Sanschagrin, S., Yergeau, E., 2014. Next-Generation Sequencing of 16S Ribosomal RNA Gene Amplicons. Journal of Visualized Experiments 3–8. https://doi.org/10.3791/51709

Schaider, L.A., Rodgers, K.M., Rudel, R.A., 2017. Review of Organic Wastewater Compound Concentrations and Removal in Onsite Wastewater Treatment Systems. Environmental Science and Technology 51, 7304–7317. https://doi.org/10.1021/acs.est.6b04778

Schirmer, M., D'Amore, R., Ijaz, U.Z., Hall, N., Quince, C., 2016. Illumina Error Profiles: Resolving Fine-Scale Variation in Metagenomic Sequencing Data. BMC Bioinformatics 17,

1–15. https://doi.org/10.1186/s12859-016-0976-y

Schloss, P.D., Jenior, M.L., Koumpouras, C.C., Westcott, S.L., Highlander, S.K., 2016. Sequencing 16s Rrna Gene Fragments Using the PacBio SMRT DNA Sequencing System. PeerJ 2016, 1–16. https://doi.org/10.7717/peerj.1869

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. Applied and Environmental Microbiology 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09

Schuchmann, K., Müller, V., 2016. Energetics and Application of Heterotrophy in Acetogenic Bacteria. Applied and Environmental Microbiology 82, 4056–4069. https://doi.org/10.1128/AEM.00882-16

Seemann, T., 2014. Prokka: Rapid Prokaryotic Genome Annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., Huttenhower, C., 2012. Metagenomic Microbial Community Profiling Using Unique Clade-Specific Marker Genes. Nat Methods 9, 811–814. https://doi.org/10.1038/nmeth.2066.Metagenomic

Sherrill-Mix, S., 2019. Taxonomizr: Functions to Work With Ncbi Accessions and Taxonomy (Version 0.5.3) [WWW Document]. URL https://cran.r-project.org/web/packages/taxonomizr/index.html (accessed 11.24.20).

Shi, C., Wang, K., Zheng, M., Liu, Y., Ma, J., Li, K., 2021. The Efficiencies and Capacities of Carbon Conversion in Fruit and Vegetable Waste Two-Phase Anaerobic Digestion: Ethanol-

529 Path vs. Butyrate-Path. Waste Management 126, 737–746.
530 https://doi.org/10.1016/j.wasman.2021.04.010
531 Shin, J., Jang, H.M., Shin, S.G., Kim, Y.M., 2019. Thermophilic Anaerobic Digestion: Effect of
532 Start-Up Strategies on Performance and Microbial Community. Science of the Total
533 Environment 687, 87–95. https://doi.org/10.1016/j.scitotenv.2019.05.428
534 Siddiqi, M.Z., Sok, W., Choi, G., Kim, S.Y., Wee, J.H., Im, W.T., 2020. Simplicispira hankyongi
535 sp. nov., a Novel Denitrifying Bacterium Isolated Grom Sludge. Antonie van Leeuwenhoek,
536 International Journal of General and Molecular Microbiology 113, 331–338.
537 https://doi.org/10.1007/s10482-019-01341-0
538 Simpson, J.T., Pop, M., 2015. The Theory and Practice of Genome Sequence Assembly. Annual
539 Review of Genomics and Human Genetics 16, 153–172. https://doi.org/10.1146/annurev-
540 genom-090314-050032
541 Sinha, R., Abu-Ali, G., Vogtmann, E., Fodor, A.A., Ren, B., Amir, A., Schwager, E., Crabtree, J.,
542 Ma, S., Abnet, C.C., Knight, R., White, O., Huttenhower, C., 2017. Assessment of Variation
543 in Microbial Community Amplicon Sequencing by the Microbiome Quality Control (MBQC)
544 Project Consortium. Nature Biotechnology 35, 1077–1086. https://doi.org/10.1038/nbt.3981
545 Sotelo, T.J., Satoh, H., Mino, T., 2019. Assessing Wastewater Management in the Developing
546 Countries of Southeast Asia: Underlining Flexibility in Appropriateness. Journal of Water
547 and Environment Technology 17, 287–301. https://doi.org/10.2965/jwet.19-006
548 Stackebrandt, E., Goebel, B.M., 1994. Taxonomic Note: A Place for DNA-DNA Reassociation
549 and 16s rRNA Sequence Analysis in the Present Species Definition in Bacteriology.
550 International Journal of Systematic Bacteriology 44, 846–849. https://doi.org/0020-
551 7713/94/$04.0O + 0

552 Steinegger, M., Mirdita, M., Söding, J., 2019. Protein-Level Assembly Increases Protein Sequence
553 Recovery From Metagenomic Samples Manyfold. Nature Methods 16, 603–606.
554 https://doi.org/10.1038/s41592-019-0437-4
555 Sun, L., Müller, B., Westerholm, M., Schnürer, A., 2014. Syntrophic Acetate Oxidation in
556 Industrial Cstr Biogas Digesters. Journal of Biotechnology.
557 https://doi.org/10.1016/j.jbiotec.2013.11.016
558 Sun, R., Zhang, L., Zhang, Z., Chen, G.-H., Jiang, F., 2017. Realizing High-Rate Sulfur Reduction
559 Under Sulfate-Rich Conditions in a Biological Sulfide Production System to Treat Metal-
560 Laden Wastewater Deficient in Organic Matter. Water Research.
561 https://doi.org/10.1016/j.watres.2017.12.039
562 Tale, V.P., Maki, J.S., Zitomer, D.H., 2015. Bioaugmentation of Overloaded Anaerobic Digesters
563 Restores Function and Archaeal Community. Water Research 70, 138–147.
564 https://doi.org/10.1016/j.watres.2014.11.037
565 Tessler, M., Neumann, J.S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L.F.M., Segovia,
566 B.T., Lansac-Toha, F.A., Lemke, M., Desalle, R., Mason, C.E., Brugler, M.R., 2017. Large-
567 Scale Differences in Microbial Biodiversity Discovery Between 16s Amplicon and Shotgun
568 Sequencing. Scientific Reports 7, 1–14. https://doi.org/10.1038/s41598-017-06665-3
569 Trivedi, U.H., Cézard, T., Bridgett, S., Montazam, A., Nichols, J., Blaxter, M., Gharbi, K., 2014.
570 Quality Control of Next-Generation Sequencing Data Without a Reference. Frontiers in
571 Genetics 5, 1–14. https://doi.org/10.3389/fgene.2014.00111
572 Truong, D.T., Franzosa, E.A., Tickle, T.L., Scholz, M., Weingart, G., Pasolli, E., Tett, A.,
573 Huttenhower, C., Segata, N., 2015. Metaphlan2 for Enhanced Metagenomic Taxonomic
574 Profiling. Nature Methods 12, 902–903. https://doi.org/10.1038/nmeth.3589

Tuan, N.N., Chang, Y.C., Yu, C.P., Huang, S.L., 2014. Multiple Approaches to Characterize the Microbial Community in a Thermophilic Anaerobic Digester Running on Swine Manure: A Case Study. Microbiological Research 169, 717–724. https://doi.org/10.1016/j.micres.2014.02.003

Tzeng, J., Lu, H., Li, W.H., 2008. Multidimensional Scaling for Large Genomic Data Sets. BMC Bioinformatics 9, 1–17. https://doi.org/10.1186/1471-2105-9-179

Vanwonterghem, I., Evans, P.N., Parks, D.H., Jensen, P.D., Woodcroft, B.J., Hugenholtz, P., Tyson, G.W., 2016a. Methylotrophic Methanogenesis Discovered in the Archaeal Phylum Verstraetearchaeota. Nature Microbiology. https://doi.org/10.1038/nmicrobiol.2016.170

Vanwonterghem, I., Jensen, P.D., Rabaey, K., Tyson, G.W., 2016b. Genome-Centric Resolution of Microbial Diversity, Metabolism and Interactions in Anaerobic Digestion. Environmental microbiology. https://doi.org/10.1111/1462-2920.13382

Venkiteshwaran, K., Milferstedt, K., Hamelin, J., Zitomer, D.H., 2016. Anaerobic Digester Bioaugmentation Influences Quasi Steady State Performance and Microbial Community. Water Research 104, 128–136. https://doi.org/10.1016/j.watres.2016.08.012

Verastigui, Y., Cheng, J., Engel, K., Kolczynski, D., Mortimer, S., Lavigne, J., Montalibet, J., Romantsov, T., Hall, M., McConkey, B., Rose, D., Tomashek, J., Scott, B., Charles, T., Neufeld, J., 2021. Multisubstrate Isotope Labeling and Metagenomic Analysis of Active. mBio. https://doi.org/10.1128/mBio.01157-14.Editor

Verma, S.K., Singh, H., Sharma, P.C., 2017. An Improved Method Suitable for Isolation of High-Quality Metagenomic Dna From Diverse Soils. 3 Biotech 7, 1–7. https://doi.org/10.1007/s13205-017-0847-x

Wang, C., Li, Y., Sun, Y., 2020. Acclimation of Acid-Tolerant Methanogenic Culture for

598 Bioaugmentation: Strategy Comparison and Microbiome Succession. ACS Omega 5, 6062–
599 6068. https://doi.org/10.1021/acsomega.9b03783
600 Wang, L., Long, F., Liao, W., Liu, H., 2020. Prediction of Anaerobic Digestion Performance and
601 Identification of Critical Operational Parameters Using Machine Learning Algorithms.
602 Bioresource Technology 298, 122495. https://doi.org/10.1016/j.biortech.2019.122495
603 Wang, P., Wang, H., Qiu, Y., Ren, L., Jiang, B., 2018. Microbial characteristics in anaerobic
604 digestion process of food waste for methane production–A review. Bioresource Technology
605 248, 29–36. https://doi.org/10.1016/j.biortech.2017.06.152
606 Wang, S., Ma, F., Ma, W., Wang, P., Zhao, G., Lu, X., 2019. Influence of Temperature on Biogas
607 Production Efficiency and Microbial Community in a Two-Phase Anaerobic Digestion
608 System. Water 11. https://doi.org/10.3390/w11010133
609 Wang, T., Zhu, G., Kuang, B., Jia, J., Liu, C., Cai, G., Li, C., 2021. Novel Insights into the
610 Anaerobic Digestion of Propionate Via Syntrophobacter fumaroxidans and Geobacter
611 sulfurreducens: Process and Mechanism. Water Research 200, 117270.
612 https://doi.org/10.1016/j.watres.2021.117270
613 Wei, S., 2016. The Application of Biotechnology on the Enhancing of Biogas Production from
614 Lignocellulosic Waste. Applied Microbiology and Biotechnology.
615 https://doi.org/10.1007/s00253-016-7926-5
616 Welch, J.L.M., Rossetti, B.J., Rieken, C.W., Dewhirst, F.E., Borisy, G.G., 2016. Biogeography of
617 a Human Oral Microbiome at the Micron Scale. Proceedings of the National Academy of
618 Sciences of the United States of America 113, E791–E800.
619 https://doi.org/10.1073/pnas.1522149113
620 Westcott, S.L., Schloss, P.D., 2015. De Novo Clustering Methods Outperform Reference-Based

Methods for Assigning 16S rRNA Gene Sequences to Operational Taxonomic Units. PeerJ 2015. https://doi.org/10.7717/peerj.1487

Westerholm, M., Dolfing, J., Schnürer, A., 2019. Growth Characteristics and Thermodynamics of Syntrophic Acetate Oxidizers. Environmental Science and Technology 53, 5512–5520. https://doi.org/10.1021/acs.est.9b00288

Westerholm, M., Moestedt, J., Schnürer, A., 2016. Biogas Production Through Syntrophic Acetate Oxidation and Deliberate Operating Strategies for Improved Digester Performance. Applied Energy. https://doi.org/10.1016/j.apenergy.2016.06.061

Wilhelm, R.C., Singh, R., Eltis, L.D., Mohn, W.W., 2019. Bacterial Contributions to Delignification and Lignocellulose Degradation in Forest Soils with Metagenomic and Quantitative Stable Isotope Probing. ISME Journal. https://doi.org/10.1038/s41396-018-0279-6

Wilkins, L.G.E., Ettinger, C.L., Jospin, G., Eisen, J.A., 2019. Metagenome-Assembled Genomes Provide New Insight Into the Microbial Diversity of Two Thermal Pools in Kamchatka, Russia. Scientific Reports 9, 1–15. https://doi.org/10.1038/s41598-019-39576-6

Wirth, R., Kovács, E., Maráti, G., Bagi, Z., Rákhely, G., Kovács, K.L., 2012. Characterization of a Biogas-Producing Microbial Community by Short-Read Next Generation Dna Sequencing. Biotechnology for Biofuels 5, 1. https://doi.org/10.1186/1754-6834-5-41

Withers, P.J.A., Jordan, P., May, L., Jarvie, H.P., Deal, N.E., 2014. Do Septic Tank Systems Pose a Hidden Threat to Water Quality? Frontiers in Ecology and the Environment 12, 123–130. https://doi.org/10.1890/130131

Wohlgemuth, R., Littlechild, J., Monti, D., Schnorr, K., van Rossum, T., Siebers, B., Menzel, P., Kublanov, I. V., Rike, A.G., Skretas, G., Szabo, Z., Peng, X., Young, M.J., 2018. Discovering

644 Novel Hydrolases From Hot Environments. Biotechnology Advances 36, 2077–2100.
645 https://doi.org/10.1016/j.biotechadv.2018.09.004
646 Wood, D.E., Salzberg, S.L., 2014. Kraken: Ultrafast Metagenomic Sequence Classification Using
647 Exact Alignments. Genome Biology. https://doi.org/https://doi.org/10.1186/gb-2014-15-3-
648 r46
649 World Health Organization, UNICEF, 2017. Progress on DrinkingWater, Sanitation and Hygiene
650 2017 Update and SDG Baselines, WHO.
651 Wright, E.S., Yilmaz, L.S., Noguera, D.R., 2012. DECIPHER, a Search-Based Approach to
652 Chimera Identification for 16S rRNA Sequences. Applied and Environmental Microbiology
653 78, 717–725. https://doi.org/10.1128/AEM.06516-11
654 Wu, Y., Wang, C., Zheng, M., Zuo, J., Wu, J., Wang, K., Yang, B., 2017. Effect of pH on ethanol-
655 type acidogenic fermentation of fruit and vegetable waste. Waste Management 60, 158–163.
656 https://doi.org/10.1016/j.wasman.2016.09.033
657 Yan, M., Treu, L., Campanaro, S., Tian, H., Zhu, X., Khoshnevisan, B., Tsapekos, P., Angelidaki,
658 I., Fotidis, I.A., 2020. Effect of Ammonia on Anaerobic Digestion of Municipal Solid Waste:
659 Inhibitory Performance, Bioaugmentation and Microbiome Functional Reconstruction.
660 Chemical Engineering Journal 401, 126159. https://doi.org/10.1016/j.cej.2020.126159
661 Yang, Z., Wang, W., Liu, C., Zhang, R., Liu, G., 2019. Mitigation of Ammonia Inhibition Through
662 Bioaugmentation With Different Microorganisms During Anaerobic Digestion: Selection of
663 Strains and Reactor Performance Evaluation. Water Research 155, 214–224.
664 https://doi.org/10.1016/j.watres.2019.02.048
665 Yenigün, O., Demirel, B., 2013. Ammonia inhibition in anaerobic digestion: A review. Process
666 Biochemistry 48, 901–911. https://doi.org/10.1016/j.procbio.2013.04.012

Yin, Q., Zhu, X., Zhan, G., Bo, T., Yang, Y., Tao, Y., He, X., Li, D., Yan, Z., 2016. Enhanced Methane Production in an Anaerobic Digestion and Microbial Electrolysis Cell Coupled System with Co-Cultivation of Geobacter and Methanosarcina. Journal of Environmental Sciences (China) 42, 210–214. https://doi.org/10.1016/j.jes.2015.07.006

Yuan, H., Zhu, N., 2016. Progress in Inhibition Mechanisms and Process Control of Intermediates and By-Products in Sewage Sludge Anaerobic Digestion. Renewable and Sustainable Energy Reviews 58, 429–438. https://doi.org/10.1016/j.rser.2015.12.261

Yue, Y., Wang, J., Wu, X., Zhang, J., Chen, Z., Kang, X., Lv, Z., 2021. The Fate of Anaerobic Syntrophy in Anaerobic Digestion Facing Propionate and Acetate Accumulation. Waste Management 124, 128–135. https://doi.org/10.1016/j.wasman.2021.01.038

Zaheer, R., Noyes, N., Ortega Polo, R., Cook, S.R., Marinier, E., Van Domselaar, G., Belk, K.E., Morley, P.S., McAllister, T.A., 2018. Impact of Sequencing Depth on the Characterization of the Microbiome and Resistome. Scientific Reports 8, 1–11. https://doi.org/10.1038/s41598-018-24280-8

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: A Fast and Accurate Illumina Paired-End Read Merger. Bioinformatics 30, 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zhang, Q., Wang, M., Ma, X., Gao, Q., Wang, T., Shi, X., Zhou, J., Zuo, J., Yang, Y., 2019. High Variations of Methanogenic Microorganisms Drive Full-Scale Anaerobic Digestion Process. Environment International 126, 543–551. https://doi.org/10.1016/j.envint.2019.03.005

Zhang, S., Chang, J., Liu, W., Pan, Y., Cui, K., Chen, X., Liang, P., Zhang, X., Wu, Q., Qiu, Y., Huang, X., 2018. A Novel Bioaugmentation Strategy to Accelerate Methanogenesis Via Adding Geobacter Sulfurreducens PCA in Anaerobic Digestion System.

https://doi.org/10.1016/j.scitotenv.2018.06.043

Zhou, X., Lu, Y., Huang, L., Zhang, Q., Wang, X., Zhu, J., 2021. Effect of PH on Volatile Fatty Acid Production and the Microbial Community During Anaerobic Digestion of Chinese Cabbage Waste. Bioresource Technology.

Zhu, W., Lomsadze, A., Borodovsky, M., 2010. Ab Initio Gene Identification in Metagenomic Sequences. Nucleic Acids Research 38, 1–15. https://doi.org/10.1093/nar/gkq275

Zhu, X., Campanaro, S., Treu, L., Seshadri, R., Ivanova, N., Kougias, P.G., Kyrpides, N., Angelidaki, I., 2020. Metabolic Dependencies Govern Microbial Syntrophies During Methanogenesis in an Anaerobic Digestion Ecosystem. Microbiome. https://doi.org/10.1186/s40168-019-0780-9

Ziels, R.M., Karlsson, A., Beck, D.A.C., Ejlertsson, J., Yekta, S.S., Bjorn, A., Stensel, H.D., Svensson, B.H., 2016. Microbial Community Adaptation Influences Long-Chain Fatty Acid Conversion During Anaerobic Codigestion of Fats, Oils, and Grease with Municipal Sludge. Water Research 103, 372–382. https://doi.org/10.1016/j.watres.2016.07.043